DOCUMENT RESUME

ED 279 859 CE 046 548

TITLE	Retirement Forecasting. Evaluation of Models Shows Need for Information on Forecast Accuracy. Volume I. Report to the Chairman, Subcommittee on Social Security and Income Maintenance Programs, Committee on Finance, United States Senate.
INSTITUTION	General Accounting Office, Washington, D.C.
REPORT NO	GAO/PEMD-87-6A
PUB DATE	Dec 86
NOTE	118p.
AVAILABLE FROM	U.S. General Accounting Office, P.O. Box 6015, Gaithersburg, MD 20877 (First five copies free; additional copies--$2.00; 100 or more--25% discount).
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Data Interpretation; Decision Making; Income; *Mathematical Models; Older Adults; *Predictive Validity; Program Costs; Research Methodology; *Research Problems; *Retirement

ABSTRACT

The Government Accounting Office (GAO) reviewed 71 actuarial, behavioral, and economic models that are used for retirement forecasting, focusing on models of federal retirement program costs, civilian retirement decisions, and retirement income. GAO wished to determine to what extent the models have been documented, to what extent the models are updated and revised, and their forecasting accuracy. Of the 71 models GAO reviewed, 32 were program cost models, 35 were retirement decision models, and 4 were retirement income models. GAO found that documented models do exist for all three retirement outcomes and that considerable effort has been made in their development and maintenance. However, model forecasts are vulnerable in several areas, including the adequacy of model documentation, the frequency or recency of model maintenance, the existence of evaluative information on model validity, and the quality of model data. With regard to documentation, GAO found that while models in all three categories have been documented, the amount, completeness, and content of the documentation varies. With regard to model maintenance or updating, this occurs regularly for program cost models, infrequently for retirement decision models, and periodically for retirement income models. Therefore, for some models, projections are based on antiquated data. With regard to validity (forecasting accuracy), GAO found that there is a serious lack of published information for most models and little evidence that serious attempts at validation are being made. The GAO concluded that Congress may wish to provide additional guidance to federal agencies responsible for the development and maintenance of retirement forecasting models. (KC)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

United States General Accounting Office

GAO

Report to the Chairman, Subcommittee on Social Security and Income Maintenance Programs, Committee on Finance
United States Senate

ED279859

December 1986

RETIREMENT FORECASTING

Evaluation of Models Shows Need for Information on Forecast Accuracy

CE046548

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

GAO/PEMD-87-6A **Volume 1**

BEST COPY AVAILABLE

GAO

United States
General Accounting Office
Washington, D.C. 20548

Program Evaluation and
Methodology Division

B-221754

December 31, 1986

The Honorable William L. Armstrong
Chairman, Subcommittee on Social Security
and Income Maintenance Programs
Committee on Finance
United States Senate

Dear Mr. Chairman:

As part of our basic continuing legislative responsibility to evaluate government programs, we initiated a review of retirement forecasting models. This report presents information we gathered on 71 models that collectively forecast three retirement outcomes: (1) the costs of federal retirement programs, (2) the retirement behavior of civilian workers, and (3) the levels and distribution of retirement income.

The report is published in two volumes. The main volume summarizes our findings across models for each of the three retirement outcomes. The supplementary volume provides technical, descriptive reviews of the individual models.

We are addressing our report to you since our findings and conclusions and the matters we suggest for congressional consideration are especially pertinent in connection with your responsibility for oversight of federal social security programs. Forecasting models are used extensively in discussions of policies for these programs.

In the report, we describe the availability of published information, or documentation, on these models, the frequency with which the models are updated and maintained for current use and the adequacy of available information on the models' potential for forecast error. We also discuss factors that influence the amount of forecast error. We show that model forecasts are vulnerable in several areas, including the adequacy of model documentation, the frequency or recency of model maintenance, the existence of evaluative information on the model, and the quality of model data. Despite these vulnerabilities, we encourage development and testing of the models and greater provision of consumer information on their quality.

We are sending copies of this report to the Secretary of Defense, the Secretary of Health and Human Services, the Secretary of Labor, the Director of the Office of Personnel Management, the Director of the Office of Management and Budget, and model developers, sponsors and other members of the modeling community. Copies will be made available to others who request them.

Sincerely,

Eleanor Chelimsky
Director

Executive Summary

Purpose

Federal outlays for retirement totaled about $200 billion in fiscal year 1984 alone and affected about 37 million recipients, forming about one third of federal domestic budget outlays. These outlays are part of a long-term trend which has accompanied the aging of the U.S. population. If, as seems likely, outlays for social security and federal worker retirement programs should continue to form a growing and difficult-to-restrain segment of the federal budget, then forecasts of retirement program costs, retirement decisions and retirement income will play an increasingly significant role in national policy. Small errors in forecasts and what can seem like minor differences among models used to generate these can have major and cumulative consequences.

Despite the importance for national policy making of sound retirement-related forecasts, information on the characteristics and quality of models used to generate projections has not been readily available. GAO therefore undertook a coordinated review of 71 actuarial, behavioral, and economic models that are used for retirement forecasting, focusing on models of federal retirement program costs, civilian retirement decisions, and retirement income. GAO asked three questions:

- To what extent have the models been documented?
- To what extent are the models updated and revised, or maintained, for future use?
- What is known about the validity (e.g. forecasting accuracy) of the models? GAO reviewed the extent to which the methods, data sources, predictors and assumptions used in the models affect forecast accuracy.

Background

A forecasting model is a mathematical representation of some aspect of reality used to predict future events, in this case, retirement outcomes. GAO examined three broad categories of retirement forecasting models. A program cost model consists of equations that include factors influencing the flow of funds into and out of Social Security or other retirement programs. A retirement decision behavior model is concerned with understanding and predicting decisions people make about working, retiring, and accepting pension benefits. A retirement income model predicts future levels and distributions of retirement income. Of the 71 models GAO reviewed, 32 were program cost models, 35 were retirement decision models and 4 were retirement income models.

Results in Brief

GAO found that documented models do exist for all three retirement outcomes, and that considerable effort has been made in their development and maintenance. However, model forecasts are vulnerable in several areas, including the adequacy of model documentation, the frequency or recency of model maintenance, the existence of evaluative information on model validity, and the quality of model data.

With regard to documentation, GAO found that while models in all three categories have been documented, the amount, completeness, and content of the documentation varies. In particular, models of program cost have been less completely documented than the other models. Retirement decision models are the most completely documented. (See pages 41-43, 65, 83.)

With regard to model maintenance, or updating and revising, this occurs regularly for program cost models, infrequently for retirement decision models, and periodically for retirement income models. However, for some models, lapses or discontinuation of essential data sets mean that projections are based on antiquated data. For example, the discontinuation of one key data set —the Longitudinal Retirement History Survey— means that most decision models must now rely upon data from 1969 regarding retirement. We already know there are more women in the present labor force than there were in 1969; there may be other differences as well in variables affecting retirement. Thus, current data are important for the predictive validity and generalizability of these models. (See pages 43-44, 65-66, 83-84.)

With regard to validity (e.g. forecasting accuracy), GAO found that there is a serious lack of published information on it for most models and little evidence that serious attempts at validation are being made. Model use rests on faith in the developers' attention to error reduction, but the user receives no documentation of validation analyses. As a result, the user cannot either select a model or interpret its results on the basis of readily available information about forecasting error. GAO found an absence of evaluations for the models reviewed, leaving questions unanswered both on the overall quality of the models and on the credibility of the modeling outcomes. (See pages 44-47, 66-71, 84-86.)

GAO's Analysis

Program Cost Models

GAO identified 32 models that forecast the cost of retirement programs in which federal employees participate. The public documentation for 29 of these consists of annual financial reports to GAO and Congress. Documentation is also available for the three models of the largest retirement programs: Military Retirement, Civil Service Retirement, and the Old Age Survivors and Disability Insurance program (Social Security). Documentation for the Civil Service and Social Security models is incomplete, but its developers indicated that they are taking steps to improve it. GAO found that cost models are the most regularly updated and maintained of the three model categories.

Published information on the validation of program cost models is available only for the military and social security models. Information is available for both on their sensitivity to changes in some assumptions and on the accuracy of some of their assumptions. GAO found no evaluations of forecast accuracy for any of the models. Forecast accuracy is influenced most by the assumptions used for predictors of costs, and also by the methods of calculation and the original data sources.

Retirement Decision Models

Documentation for the 35 retirement decision behavior models consists of one or more research papers or professional journal articles, and is focused largely on theoretical aspects of each model. For these models, there is little individual updating or maintenance. Rather, it is more common for developers to construct new models than to update older ones. The models are based on restricted subgroups of the population—largely men—and outdated information. The decreasing availability of current sources of nationally representative longitudinal data poses future maintenance and generalizability problems for these models.

The publication of information on model validation is irregular. Theoretical validity is reported the most consistently and voluminously. Data validity, however, is largely untreated. Information on the models' ability to explain past retirement behavior is poor, and tests of their ability to predict future behavior have not been reported. Factors that affect forecast accuracy are methods of estimation, selection of core and other assumptions, predictors of retirement behavior and data sources.

Retirement Income Models

The 4 models of retirement income have been well documented, and the documentation has been revised to reflect retirement policy changes in the Social Security program through 1983. Income models are not updated and revised routinely, although some revisions to them have been made recently and are being made periodically.

The validation of retirement income models is poor, or at least, poorly documented, in contrast to the general documentation of these models. Even summary validation procedures are not often documented. Forecast error in income models can arise from methods of estimation, data sources, and selection of predictors and their values. Of the three types of models, these are the most speculative. The great opportunity for error suggests caution in interpreting their forecasts.

GAO believes that despite their vulnerabilities, the models are useful for a variety of purposes, especially analyzing the effects of public policy changes. Therefore, GAO believes that further development and testing of the models is appropriate. In particular more validation and documentation of these models are needed, which should, in turn, result in a greater provision of consumer information on the quality of forecasting models used for retirement policy-making.

Recommendations

GAO makes no recommendations.

Matters for Congressional Consideration

Congress may wish to provide additional guidance to federal agencies responsible for the development and maintenance of retirement forecasting models. In particular, more systematic information is needed on how developers validate their models, what the results of those validation efforts are and what they mean with respect to potential forecast error.

Agency Comments

GAO asked the Departments of Defense, Health and Human Services, and Labor and the Office of Personnel Management to review and comment on a draft of this report. Only HHS disagreed with our conclusions on the weakness of most model documentation, and we revised our description of one HHS model based on their comments. On model maintenance, HHS and DOL noted that although our text was not incorrect, decision models serve different purposes and are therefore updated less than cost models. All but OPM commented on assessing models' validity and suggested alternatives, but after fully considering their comments we did not change our conclusions about the importance of forecast accuracy as a key criterion or about the difficulty of evaluating models that did not report data on accuracy.

Contents

Executive Summary 4

Chapter 1 Introduction 14
What Is a Retirement Forecasting Model? 14
Previous Work and the Need for the Present Study 15
Objectives 16
Scope and Methodology 17
Strengths and Limitations 23
The Organization of This Report 23

Chapter 2 Models of Federal Retirement Program Costs 24
Background and Use 24
Descriptive Dimensions 27
Analytic Dimensions 41
Summary 47

Chapter 3 Models of Retirement Decision Behavior 50
Background and Use 52
Descriptive Dimensions 52
Analytic Dimensions 65
Summary 71

Chapter 4 Models of Retirement Income 74
Background and Use 74
Descriptive Dimensions 75
Analytic Dimensions 83
Summary 86

Chapter 5 Conclusions, Matters for Consideration, Agency Comments and GAO's Response 88
Summary 88
Conclusions 90
Matters for Consideration by the Congress 91
Agency Comments and GAO's Response 91

Appendixes

Appendix I: Comments From the Department of Defense 98

Appendix II: Comments From the Department of Health and Human Services 100
Appendix III: Comments From the Department of Labor 107
Appendix IV: Comments From the Office of Personnel Management 114

References

115

Glossary

118

Tables

Table 2.1: Total Participants and Forecast Date For Retirement Programs Reporting Under P.L. 95-595 and OASDI 26
Table 2.2: Actuarial Cost Methods Used in Models of Federal Retirement Programs 30
Table 2.3: Economic Assumptions Used in Models of Federal Retirement Programs Reporting Under P.L. 95-595 for the 1983 Plan Year 36
Table 2.4: Alternative II-B Economic Assumptions Used in the 1983 OASDI Forecast 39
Table 2.5: Average Difference Between Actual and Forecast Values for Selected Economic Assumptions for OASDI Trustee Report Years 1973-81 46
Table 3.1: Models of Retirement Behavior 51
Table 3.2: Data Sources for Retirement Behavior Models 57
Table 3.3: Model Treatment of Sex Differences 59
Table 3.4: Model Treatment of Race Differences 60
Table 3.5: Model Treatment of Marital Status Differences 60
Table 3.6: Model Treatment of Social Security Effects 62
Table 3.7: Predictors Used in Retirement Decision Models 64
Table 4.1: Available Income Breakdown for Models of Retirement Income 77
Table 4.2: Available Demographic Breakdown for Models of Retirement Income 79
Table 4.3: Comparison of Projections From DYNASIM and PRISM 85

Abbreviations

AARP	American Association of Retired Persons
CPS	Current Population Survey
CSRS	Civil Service Retirement System
DI	Disability Insurance
DOD	U.S. Department of Defense
DOL	U.S. Department of Labor
DRI	Data Resources, Inc.
DYNASIM	Dynamic Simulation of Income Model
ERISA	Employee Retirement Income Security Act of 1974
FICA	Federal Insurance Contributions Act
GAO	U.S. General Accounting Office
HHS	U.S. Department of Health and Human Services
MALTHUS	Michigan Quarterly Econometric Model of the U.S.
MDM	Macroeconomic-Demographic Model
NLS	National Longitudinal Surveys of Labor Market Experience
OASI	Old Age and Survivors Insurance
OASDI	Old Age Survivors and Disability Insurance
OMB	Office of Management and Budget
OPM	Office of Personnel Management
PBGC	Pension Benefit Guarantee Corporation
PRISM	Pension and Retirement Income Simulation Model
PSID	Panel Study of Income Dynamics
RHS	Retirement History Survey
SER	Summary Earnings Record
SSI	Supplemental Security Income
STRS	State Teachers Retirement System
TRIM	Transfer Income Model

Chapter 1

Introduction

In January 1983, the bipartisan National Commission on Social Security Reform endorsed a controversial package of proposals aimed at rescuing the Social Security trust fund from imminent depletion. Critical information for the Commission's decision-making included forecasts from the Social Security cost model, which both forecasted the magnitude of the program's financing problem and evaluated the savings in costs to be expected from each of the various proposals the Commission considered. The Commission's proposals, which were soon enacted with only minor modification despite the political sensitivity of the issues, provide one example of how retirement-related forecasting models can be of use in public policy analysis and decision-making.

This report reviews 71 models, including 32 models of retirement program costs, 35 models of retirement decision behavior, and 4 models projecting retirement income. The remainder of this chapter provides our definition of a retirement forecasting model, a summary of previous studies on retirement models and their implications for the present report, and our objectives, scope, and methodology. The chapter concludes with a summary of the organization of the remainder of the report.

What Is a Retirement Forecasting Model?

In this report, we define a retirement forecasting model as a mathematical representation of some aspect of reality used to predict future events—in this case, retirement outcomes—given the present situation, or to determine the likely consequences of changes in the present on future events. We thus include models which project the future as well as models that are used primarily to analyze the future consequences of policy change.

Some models we discuss as a result of adopting this definition may not be routinely referred to as forecasting models.[1] One example is the combination of actuarial methods and formulae which are used to project the future cost of a retirement income program. These equations, which express the relationship among the factors that influence program costs, meet the definition of a model and thus are included in our review.

---

[1] Models developed primarily for the purpose of explaining or describing the present could be used to make predictions and are thus potential forecasting models, although they typically are not referred to in that way. This report reviews models which actually produce forecasts as well as models which could produce them.

Previous Work and the Need for the Present Study

GAO has a long history of involvement with modeling issues. This history, through 1978, was summarized in Models and Their Role in GAO (GAO, 1978a). GAO reports in the area have included inventories of models, evaluations of specific models and of model uses, and general guidelines and recommendations for evaluating models. GAO also uses models developed by others, such as the large macroeconomic models of the national economy and on occasion, develops its own models. The present report is the first GAO summary of models in the retirement area.

Several prior studies have identified some retirement-related models but present various limitations—such as omitting models, being out-of-date, and providing little descriptive information—that this report proposes to remedy. We found no single source which identifies models for all three purposes—forecasting program costs, retirement behavior and retirement income—which the present report does.

An extensive literature search yielded few descriptive or comparative reviews of models for policy-makers or other model or forecast users. Exceptions include reviews of the Old Age Survivors and Disability Insurance (OASDI)[2] cost estimate model by GAO (1983b, 1986) and others (e.g. Myers, 1982; Light, 1983), several GAO reviews of individual federal retirement programs (e.g. 1982d, 1983a, 1985), and a comparison of two computerized models of retirement income (Haveman and Lacker, 1984). The content of these reviews is varied and because many models are constantly being updated and revised, earlier reviews are not always current or relevant.

Retirement forecasting models can play an important role in policy debates over new and existing retirement policies and programs. For example, the virtually universal coverage of workers by the OASDI program makes this program of central concern to the public, and as a consequence, it has been the focus of many modeling activities. The Office of the Actuary at the Social Security Administration develops and maintains forecasting models in order to monitor the financial status of the OASDI program as well as the financial impact (or costs) of proposed changes to that program. Also, many models of retirement decision behavior have been developed to assess the nature of any work disincentive effects in the OASDI program and to forecast the effects of proposed policy changes on the future work and retirement decisions of individuals. Finally, retirement income models have been used to forecast the amount of retirement income that will be available from the OASDI program and its likely distribution across various subgroups of the population. Although the OASDI program has been the focus of much modeling activity, models of other retirement programs have also been used for public policy analysis.

---

[2] The OASDI program is a public insurance program administered by the Social Security Administration. The term "social security" is frequently used to describe a combination of programs, including Old Age and Survivors Insurance (OASI), Disability Insurance (DI), Hospital Insurance (HI) and Supplementary Medical Insurance (SMI). In this study, the term is used interchangeably with OASDI, the retirement and disability benefit components of the social security program.

The bewildering array of models, the lack of a current inventory of models from actuarial, behavior and economic disciplines, and the importance of these models in public policy making suggested the need for the present study.

Objectives

This report is a guide for both users and policy makers to models of three retirement outcomes—retirement program costs, retirement decision behavior, and retirement income. The three outcomes were chosen because of the Congressional interest in ensuring that future retirement benefits are soundly funded, in monitoring future labor supply and future rates of application for retirement benefits, and in promoting equity in the distribution of retirement income and a minimum income level for the elderly retired. Models of all three outcomes were used in the social security reform debate which culminated in the enactment of Public Law 98-21, the 1983 Social Security Act amendments.

Our intent is to provide a guide to models of each of these three outcomes in the form of model reviews and individual descriptions of the models identified for each outcome. One issue of central concern to forecast users and policymakers is forecast error—the extent to which actual experience differs from forecasts. Our model reviews provide information on likely sources of error in these models.

Scope and Methodology

Identifying Models

Models of Retirement Program Costs

As an employer, the federal government is not only concerned with the retirement income security of its workers but with the cost to the government of providing benefits to federal employees. Thus, models which forecast future costs of federal retirement programs are important for short-term budgeting decisions and long-term decisions about changes in program structure that may be needed to ensure continued ability to pay benefits to retirees. We identified 32 cost models for each of 37 retirement income programs that are administered by the federal government and provide benefits to federal workers. One of these is the OASDI cost estimate model. The others were identified by reviewing the most recent annual reports of federal pension plan administrators.[3]

Three types of retirement cost models are not covered in our review. First, our review does not include models of the costs of private or state and local government pension plans. The federal government maintains oversight of the financial solvency of private pension plans primarily through the Employee Retirement Income Security Act (ERISA), which requires plan sponsors to disclose financial information, and through the establishment of the Pension Benefit Guarantee Corporation (PBGC) which insures benefits against plan terminations. There are over 600,000 private pension plans and hundreds of state and local government plans and each potentially has an associated model for forecasting future costs of the program. In addition, models maintained by the PBGC to forecast plan terminations are not reviewed here.

Second, we do not review models associated solely with disability benefits, such as those provided through the Veterans Administration, models associated with the Railroad Retirement System, which is managed by the federal government but covers private sector workers, and models associated with private retirement programs that cover some federal employees but primarily insure private sector workers and are thus monitored under ERISA.

---

[3] For a summary of these reports, see Summary of 1983 Federal Pension Plan Information, GAO/AFMD-85-69.

Third, we exclude budgetary models of annual or quarterly retirement program costs. These models are developed by federal agencies and are used internally by them to project quarterly and annual budget needs. The Congressional Budget Office, the Office of Personnel Management and the Department of Defense are examples of groups that sponsor such models.

Models of Retirement Decision Behavior

We identified models that predict or explain the decision of workers to retire through an initial literature search that was supplemented by surveying model developers and other experts in the field. The initial search yielded 42 potentially relevant models developed by 28 research teams. A request letter which solicited additional model identification was mailed to developers and experts. Seventy-nine percent of those contacted responded by mail or phone and an additional 43 potential models were identified. We excluded from review (1) models of retirement plans or intentions, (2) models of aggregate retirement trends, (3) theoretical models of retirement behavior that have not been empirically estimated, (4) models of military retirement behavior, and (5) models in unpublished doctoral dissertations. By applying these criteria, we identified a final set of 35 models.

These 35 models all specify a set of factors that are hypothesized to influence workers' decisions to retire and all test a theoretical model on actual observations of behavior (as opposed to stated plans or intentions) from surveys of individuals or administrative records of employees. These models examine the effects of both private and public pension income on retirement behavior. Some are based only on private sector employees, some only on public sector employees and others on both.

Models of Retirement Income

The final category focuses on large scale computerized models which are designed for making long-range forecasts of the levels and/or distribution of retirement income. We identified four of these through an initial literature search supplemented by contacting experts in the field and interviewing executive agency personnel who were the most likely users and sponsors of such models. These models identify multiple sources of income such as social security, private pension plans, personal savings and investment, and the like, in order to consider total income available to retirees. It is total income that will provide the fullest indicator of potential standard of living. In view of the limited number of available models, we also included comprehensive models of the income of elderly persons noting that not all elderly are retired and vice versa. We excluded from review: (1) macroeconomic models of the U.S. economy, (2) single equation models, (3) purely theoretical models, and (4) models of single sources of retirement income (such as IRAs). These have narrower applications for retirement policy.

Describing and Reviewing Models

The extent of information on individual models presented in this report is greater than in previous GAO inventories (e.g., GAO, 1979, 1982a) but less than the information that could be provided in an in-depth evaluation of a single model (e.g., GAO, 1977b).

We established a general framework for reviewing all models by selecting descriptive and analytic dimensions relevant to all model categories. The selected descriptive dimensions include:

- Outcomes—primary model outcomes
- Methods—mathematical technique/method used
- Data Sources—primary external sources of data
- Predictors—factors that influence outcomes and how their values were derived

The analytic dimensions include:

- Documentation—availability of user-oriented documentation and its contents
- Maintenance—frequency of model updating and revision
- Validity—procedures used by model developers to monitor the divergence between real world and model outcomes

Using Ascher's terminology, the descriptive dimensions provide information on the formulation of a forecast from an "insider's" or forecast specialist's point of view. The analytic dimensions provide information from a perspective that is "outside" the forecasting endeavor (Ascher, 1978, p. 7). Both perspectives are important for evaluating the credibility of a forecast.

Descriptive Dimensions

The four descriptive dimensions and additional background information included in the review were developed by aggregating from a checklist of 42 information items that are recommended for inclusion in model documentation (McLeod, 1973). Items in McLeod's checklist that did not generalize across model categories, such as computer running time, were deleted for the present review. We excluded other items on the checklist, such as simulation results and justification of assumptions, because all of the models have numerous results and assumptions so that fully describing them in detail would not be feasible for a review of 71 models. Where possible in our general analyses and individual model reviews, we deal with key assumptions. The selected set of dimensions provide basic information on model development. In addition, we identify sources (e.g. model developer or sponsor and a document) where the reader can obtain more detailed information.

The outcomes dimension refers to specific outcomes that the models produce. For example, retiring from work and drawing a pension benefit are two outcomes that could be predicted by a retirement decision behavior model. The outcomes dimension is probably the most important for allowing model users to determine which models are relevant for their purposes.

The methods dimension refers to the actual techniques used to implement the models reviewed. Most of these models have been derived with a series of statistical analysis and simulation techniques. These techniques vary from model to model, and have important effects on forecasted outcomes.

Data sources supply the basic information, obtained externally, that the model processes in generating its forecasts. For example, a model may predict retirement income by taking as one data source information collected every year by the Bureau of the Census. The accuracy and reliability of that data are important to the overall credibility of forecasted outcomes. The source of data also determines what kind of population the model depicts and thus the generalizability of results.

The predictors are a set of factors used to describe different aspects of the system being modeled. While the three previous dimensions remain constant when a forecast is generated, the values for the predictors can vary and this variation produces variation in the outcomes for different individuals or groups.[4] For some models the set of predictors is specified in the choice of methods. For others, the set of predictors is selected from a combination of theory and historical observations of relationships between factors and the modeled outcome. Depending on this mix, the nature of the relationship between outcomes and the predictors can be either specified or estimated.

---

[4] Variation in outcomes for different individuals or groups is not completely explained by variation in the predictors. The unexplained variation is a component of the forecast error.

Predictor values can be based on actual observations of individuals or groups (e.g., the assets or income of an individual or the total amount of retirement benefits paid to a group of individuals). Or, they can be based on assumptions about future values. In the case of retirement cost models, these assumptions are usually classified as demographic (e.g., size of population) or economic (e.g., inflation rate). These numbers are sometimes taken from other forecasting models, are sometimes derived through expert judgment, are sometimes estimated from historical observations or in other cases are explicitly controlled. The core assumptions underlying a forecast are major determinants of forecast accuracy. Of particular importance is the avoidance of "assumption drag"—reliance on outdated assumptions (Ascher, 1978, pp. 199-203).

Other Descriptive Information

This report also presents general background information for all models including identification of a model (name, history, developer), model purpose, how the model has been used, what provisions are made for revising and updating the model, and the potential for future use of the model.

Analytic Dimensions

In developing analytic criteria for reviewing models, we relied upon our publication, Guidelines for Model Evaluation (GAO, 1979) and on standards for evaluating models and appraising forecasts recommended by others (e.g., Gass, 1976; Ascher, 1978; Anderson, 1980).

Guidelines for Model Evaluation proposes five primary criteria for evaluating models: documentation, maintainability, validity, computer model verification, and usability.[5] Of these five criteria, three were selected for the present study: documentation (written general information about a model), maintainability (the extent of model review and updating) and validity. Computer model verification and usability require hands-on use of the models, an activity that is beyond the scope of this review. Within the validity dimension, only operational validity (the extent of divergence between the "actual" and the outcomes predicted by the model) is included as a review dimension. Operational validity includes forecast accuracy. Although there was in the past some lack of agreement among forecasting specialists over the use of an accuracy criterion in the appraisal of forecasts, more recent debate reflects considerable consensus in the field that accuracy and other indicators of operational validity unquestionably increase the value of a forecast for policy decision-making. Data and theoretical validity are also of central importance in assessing the credibility of forecasts. However, the multiple data sources used in the models we reviewed and the complexity of the underlying model theories made it impossible to review these in depth for each individual model. However, we address general data validity and theoretical validity issues in our general analyses of the broad groups of models we review.

---

[5] The importance of each of these criteria, especially to decision makers who rely on the results of modeling efforts, is discussed in detail in our 1979 publication.

In the present inventory, we discuss the documentation, maintainability (which we renamed maintenance for this review) and validity dimensions and cite the kinds of information available on each of these dimensions. Although we searched for complete evaluations of models based on the standards mentioned above, we found none for the models we reviewed.

Information Sources

Most information in this report came from publicly available documentation. Additional information was obtained from reviews of individual models, identified through a literature review, and from interviews with model developers, users, and experts in the field.

The sources varied across categories. Documentation for the models of retirement program costs was obtained primarily from reports submitted annually to GAO from federally administered retirement programs. Additional documentation for the Department of Defense model of the Military Retirement System, the Office of Personnel Management model of the Civil Service Retirement System and the Social Security Administration model of the OASDI program came directly from the agency developers.

Documentation for models of retirement decision behavior consists largely of the article or paper in which the model (usually one or two equations) is described. We acquired these documents through libraries or direct request to authors.

Documentation for retirement income models was obtained directly from the model developers. Although there are a limited number of models in this category, the documentation for these complex models is fairly extensive.

Strengths and Limitations

A key strength of this report is accuracy of description. In addition to our own checks, we provided model developers with our descriptions of their models and invited their review for accuracy. Ninety-nine percent (70 of 71) of the developers responded and all identified errors were corrected. The individual model descriptions are contained in the supplementary volume of this report. Accuracy of our description does not imply that the models themselves are accurate.

Three limitations may be important to mention. First, for the model whose developer did not respond to our request for review, completeness and accuracy is limited to the extent that the published documentation is limited. Second, as we noted earlier, our analyses of the models in this report are not in-depth model evaluations and therefore are not definitive statements on individual model quality. That is, we did not verify accuracy of coding, conduct hands-on tests of the programming, or test data validity. Thus, we refer to the reports on individual models as "descriptive reviews." Third, our data collection was completed in December 1984 and it is likely that some new models have been developed in one or more of the categories, or at least that changes have occurred in existing models since that time.

The Organization of This Report

Chapters 2, 3 and 4 present our overall reviews of cost, decision and income models, respectively. Chapter 5 provides an overall summary of the report and our conclusions, matters for Congressional consideration, agency comments and our response. Copies of the agency comments are in appendices. A reference list provides full citations for publications mentioned in the text. The individual model descriptive reviews are provided in the supplementary volume of this report, which also contains references for each model and a complete bibliography of literature reviewed in preparing the main and supplementary volumes.

Chapter 2

Models of Federal Retirement Program Costs

In this chapter we review 32 models which forecast the expected cost and financial status of retirement programs that cover federal employees. Public Law 95-595, the 1978 amendment to the Budget and Accounting Procedures Act of 1950, now codified at 31 U.S.C. §§ 9501-9504 (1982), requires all sponsors of federal retirement programs not covered under ERISA[1] to report annually on their financial status. Among these programs are the Civil Service Retirement System (CSRS) and the Military Retirement System as well as 29 additional programs whose sponsors annually make forecasts of their financial status.[2] In addition to models of these programs, we include the model of the OASDI program which also covers some federal employees.[3] The supplementary volume of this report contains a descriptive review of each of these agency sponsored models.

These forecasting models have many features in common, including similar outcomes and predictors. They differ in the choice of values for their predictors (i.e., their assumptions about future economic and demographic trends), in the methods used to calculate the future financial status and costs of the program, and finally, in the specific characteristics of each pension plan and the number of participants in each plan.

We describe these models across the descriptive and analytic dimensions presented in chapter 1, discussing their similarities and differences. Throughout this chapter we first discuss the 31 models developed for programs reporting under P.L. 95-595 and then contrast them with the OASDI model.

Background and Use

The models described here (except for the OASDI model) are used by each plan sponsor to produce annual P.L. 95-595 reports. Public Law 95-595 is designed to protect the interests of participants in federal government pension plans by requiring plans to report annually on their financial condition. These reports are similar to those required of private plans under ERISA. The specifics are determined by the President in conjunction with the Comptroller General of the United States; responsibility has been delegated by the President to the Office of Management and Budget (OMB). There are explicit instructions on how the financial information should be calculated and presented and how the modeling methods should be described. These reports provide general information on the plans and their methods and were the primary basis for our model reviews.

---

[1] The Employee Retirement Income Security Act of 1974.

[2] There are a total of 46 retirement plans covered by P.L. 95-595 (see Summary of 1983 Federal Pension Plan Information, GAO/AFMD-85-69). Six of these are defined contribution plans and thus do not have associated forecasting models: three TIAA/CREF-administered plans (Smithsonian, Uniformed Services University and Department of Agriculture) and three others (Pearl Harbor Restaurant, Spokane Production Credit Association and Spokane Thrift). No reports had been filed at the time of our review for two plans: the President's Retirement plan and the Federal Home Loan Bank Pension Portability plan. (The President's Retirement plan has since filed reports for 1984 and 1985.) We omitted from review two additional plans: the Comptroller General's Retirement plan and the Army Stars and Stripes plan. The former plan has no active participants and the latter at the time of our review had not filed a report since 1980 and was being phased out. For the remaining 36 plans, we identified 31 models; three of these forecasted outcomes for two plans each and one, for three plans.

[3] The Social Security Administration has more than one forecasting model for the OASI and DI programs. We refer to the set of models as a single OASDI cost estimate model.

The OASDI cost model is used primarily to generate projections for the Annual Report of the Board of Trustees of the OASI and DI Trust Funds, and to provide predictions of the results of potential program policy changes. The Office of the Actuary is the developer of the models and has been making forecasts for the programs for close to 50 years.[4]

The complexity of individual models of these retirement income programs varies from a set of equations and static, or stationary, assumptions about future economic and demographic trends applied to a baseline population, to a series of sub-models which apply dynamic future assumptions (ones that change over time) to simulated future populations. The degree of model complexity is related to the size of the program's covered population, with the OASDI program being by far the largest and having the most complex forecasting model associated with it.

A summary of program size for the 31 models which currently report under P.L. 95-595 as well as the OASDI program appears in table 2.1. As table 2.1 illustrates, the plans vary widely in size. The OASDI program, which covers most of the U.S. adult population, is the largest. Among the programs which report under P.L. 95-595, the CSRS and Military Retirement System, which each have over four million participants, are by far the largest. The Tax Court Judges System (no. 31 in table 2.1) with 28 participants is the smallest. Overall, there are over 9.5 million participants in these P.L. 95-595 programs with approximately 6.0 million active employees, 2.8 million retirees, 0.6 million other (disability and survivor) beneficiaries, and 0.2 million separated employees with benefit rights.

---

[4] Although forecasts have been generated by the Office of the Actuary for 50 years, it is inappropriate to view estimates from the OASDI model as coming from a single model with a 50-year history because the procedures used to derive the final estimates have changed substantially across this time period.

Table 2.1: Total Participants and Forecast Date[a] for Retirement Programs Reporting Under P.L. 95-595 and OASDI

Retirement system[b]	Total participants[c]	Forecast date
1. Civil Service	4,754,000	9/30/83
2. Military	4,533,195	9/30/83
3. Coast Guard	73,913	12/31/83
4. Federal Reserve	35,402	12/31/83
5. Army/Air Force[d]	32,943	12/31/83
6. Tennessee Valley Authority	32,833	9/30/83
7. Navy/Coast Guard Resale[e]	18,973	12/31/83
8. Foreign Service	19,553	9/30/83
9. Army	13,224	9/30/83
10. Public Health Service	7,806	9/30/83
11. Air Force	6,494	9/30/83
12. Marines	4,253	12/31/83
13. Louisville, KY FCB[f]	2,931	12/31/83
14. St. Paul, MN FCB	2,820	12/31/83
15. Omaha, NE FCB	2,462	12/31/83
16. Columbia, SC FCB	1,977	8/31/83
17. St. Louis, MO FCB	1,957	12/31/83
18. Wichita, KN FCB	1,371	2/28/83
19. Sacramento, CA FCB	1,321	12/31/83
20. Spokane, WA FCB	1,342	12/31/83
21. Judiciary[g]	1,117	12/31/83
22. Baltimore, MD FCB	1,048	12/31/83
23. Austin, TX FCB	1,001	12/31/83
24. Springfield, MO FCB	908	3/31/83
25. Jackson, MI FCB[h]	630	12/31/83
26. Jackson, MI FCB, Production Credit Association[h]	607	12/31/83
27. National Oceanic and Atmospheric Administration	542	9/30/83
28. Federal Home Loan Mortgage	536	12/31/83

Retirement system[b]	Total participants[c]	Forecast date
29. Norfolk Naval Shipyard	108	12/31/83
30. Navy Morale, Welfare and Recreation	85	12/31/83
31. Tax Court[i]	28	12/31/83
Total	**9,555,380**	
32. OASDI[j]	115,222,000	covered workers
	35,811,000	beneficiaries

[a] The dates given here are those of the forecasts on which our reviews are based.

[b] The system name is abbreviated. The full name for each plan is provided in the supplementary volume to this report.

[c] The total of individuals who are working and covered by the plan, individuals no longer working who are entitled to benefits, and individuals receiving benefits.

[d] This system includes two plans: the annuity plan for Army/Air Force Exchange Service employees and a supplemental plan for members of the Executive Management Program.

[e] This system includes three plans: the Navy Resale, Navy Personnel and Coast Guard Resale plans.

[f] FCB denotes the retirement plan for Farm Credit Banks in the district represented by the city listed.

[g] This system includes two plans: the Judiciary and Judiciary Survivors plans.

[h] Jackson is the new location for the 5th FCB District, formerly centered in New Orleans.

[i] This system includes two plans: the Tax Court and Tax Court Survivors plans.

[j] These figures are estimates for the 1983 calendar year, based on Alternative II-B assumptions, as given in the Board of Trustees 1983 Annual Report, p. 75.

Table 2.1 also includes the effective date of the forecast on which our reviews are based. For the majority of plans, that was for the end of the 1983 plan year. This was the most recent date for which data from all plans was available at the time of our review. The models use different valuation dates because they define the plan year differently. Most plans use either a fiscal or calendar year forecast cycle.

In the following sections we will first describe cost models across the four descriptive dimensions and then review the models in terms of their documentation, maintenance and validity.

Descriptive Dimensions

From among the four descriptive dimensions outlined in chapter 1 of this report (outcomes, methods, data sources, and predictors), two are the most important for describing how the cost estimate models reviewed here differ: methods and predictor values. We describe the typical outcomes and the extent to which they vary across models. The methods used by the models to estimate the outcomes do vary and are important for understanding model forecasts. The data sources for the models are administrative records on the characteristics of plan participants and external sources which provide values for predictors. The predictors used by the models do not differ significantly among models, but the values of those predictors—that is, the economic and demographic assumptions—do differ significantly across the models and are among the most important determinants of the model outcomes.

Outcomes

The outcomes of cost estimate models are of central use in ensuring that plans are properly funded. The plans for many of the models described here are funded through regular contributions by the federal government employer and/or employees although there are several possible funding strategies. The idea is to fund the plan properly so that there will be sufficient resources to pay off current and future beneficiaries. Balancing the inflows (contributions and asset earnings) against the outflows (benefits paid and administrative expenses) would be a relatively simple matter if all inflows and outflows were made in the same year, but inflows and outflows from a fund occur over a long time period. Some employees working and contributing in the current year may not receive their benefits until 30 years from now, yet their current contributions and benefits (and contributions and benefits for all other employees and beneficiaries) must be taken into account in balancing the flow of funds. The outcomes of cost estimate models do precisely that.

The P.L. 95-595 models generate outcomes using standard private pension plan valuation methods. An actuarial valuation is the determination, as of the valuation date, of the normal cost, actuarial liability, unfunded actuarial liability, value of assets, and related present values for a pension plan. These outcomes are used to determine the financial status and cost of a pension plan, and are defined in the glossary. From the employer's viewpoint, the normal cost and the payment for the unfunded actuarial liability represent the employer's annual pension expense and hence valuation models can be viewed as cost models.

There is no single "correct" normal cost or actuarial liability for a given pension plan in a given year. Rather there are a variety of correct costs and liabilities that are determined by the methods and assumptions used in the valuation. The determination of these outcomes does have a predictive element, however, because it is necessary to make predictions on what future expenses and revenues will be.

The annual pension expense can be divided into two parts: the normal cost and an additional contribution to pay off the unfunded actuarial liability. An unfunded actuarial liability can arise for several reasons and it is normal for plans to have such a liability. Most plans start up with an actuarial liability because employees are granted credit for past service but no contributions have yet been made for benefits which will arise from that service. The actuarial liability can change from year to year for several reasons. Differences between reality and expectations for the actuarial assumptions (predictor values) can cause the actuarial liability to grow or shrink, depending on the direction of the error. In addition, the actuarial liability can change if changes are made to the plan rules. Changes in the formula for calculating benefits can affect the actuarial liability because previous contributions to the plan were based on the old benefit formula and the contributions may not have been sufficient in light of the new benefit formula. Although not all employers set aside a portion of money to pay beneficiaries, the normal cost is still calculated because it provides information on the theoretical cost of the plan if it were funded and is a basis for comparing plans.

Because of differences in the structure of the OASDI program relative to other pension programs, the outcomes forecasted by the OASDI cost estimate model are different. In contrast with the P.L. 95-595 models, the OASDI cost estimate model explicitly forecasts expenses and revenues for 75 years into the future. The outcomes for the OASDI model are presented as a percentage of taxable payroll, instead of as dollar figures. The revenue rate is essentially the social security tax rate which has been legislated for a particular year.[5] The expense rate is based on a forecast of benefits to be paid in that year (expressed as a percentage of taxable payroll). Normal cost and the unfunded actuarial liability are not relevant for the OASDI model, but the model does calculate average cost which is the average of the expense rates over the 75-year period. The average cost indicates the recommended average taxation rate over the 75-year forecast horizon that would be required in order for the program to be in actuarial balance.

Methods

The particular actuarial cost method used to produce the forecast determines how normal cost (or average cost for the OASDI model) and the unfunded actuarial liability are calculated for a plan with a given set of assumptions. A variety of methods are used to make an actuarial valuation.[6] For reporting purposes, all models produce a closed group valuation: that is, costs are calculated only for current plan participants. Some models also produce an open group valuation: new entrants are figured into the cost calculations. The OASDI, CSRS, and Military Retirement models all perform open group as well as closed group valuations.

---

[5] The rate is not simply the sum of the employer and employee rates legislated for a particular year. It is necessary to adjust these rates because of the way covered payroll is defined in the calculations.

In addition to the open/closed group distinction, the models use either a balance sheet method (forecasts are presented as of a certain time period with information about future expenses and revenues aggregated in present value calculations) or a projection method (forecasts of expenses and revenues are made and reported explicitly for years into the future). Of the models reviewed here, only the OASDI cost estimates use a projection method in their basic valuation. The remainder use balance sheet methods. The balance sheet methods can additionally be classified across three dimensions: the treatment of the benefit, the treatment of the unfunded actuarial liability, and the level of analysis. These dimensions and specific methods within each are shown in table 2.2. The number of models using each method is also indicated there.

Table 2.2: Actuarial Cost Methods Used in Models of Federal Retirement Programs[a]

	Benefit treatment	
Level of analysis	**Projected benefits**	**Accrued benefits**
Individual		
With Actuarial Liability	Entry age normal (17)	Unit Credit (2)
Without Actuarial Liability	Individual level premium (1)	Not Applicable
Aggregate		
With Actuarial Liability	Aggregate entry age normal (3)	Not Applicable
	Frozen initial liability (8)	
Without Actuarial Liability	Aggregate (2)	Not Applicable

[a]The number of models using each method is indicated in parentheses. Two plans used more than one method so there is a total of 33 method applications for 31 models.

There are two approaches to the treatment of benefits. The accrued benefit approach calculates normal cost by taking the present value of the portion of benefits earned in the year of the valuation. Projected benefit approaches consider the present value of all benefits including those already earned and yet to be earned in the calculation of normal cost. Table 2.2 shows that 31 of the 33 method applications are of the projected benefit type.

---

[6] A valuation for the purpose of this discussion refers to any forecast generated by a cost model, including the OASDI cost estimate model, although it does not technically meet the original valuation definition because it generates different outcomes.

There are also two approaches to the treatment of the unfunded actuarial liability. When the accrued unfunded liability is included as a part of the normal cost calculation, the approach is called a without actuarial liability method. Forecast methods which do not include the accrued unfunded liability in the normal cost calculation are referred to as with actuarial liability methods. The with and without actuarial liability distinction is only relevant for projected benefit methods since an accrued benefit actuarial cost method will have an actuarial liability separate from the normal cost by definition. Thirty of the 33 method applications are of the with actuarial liability type.

Finally, there are two levels of analysis: individual level and aggregate level. This distinction again is only relevant for projected benefit methods, since accrued benefit methods are all based on the individual level of analysis. Table 2.2 shows that 20 of the 33 method applications are at the individual level of analysis.

Balance sheet methods differ in ways other than on the three dimensions depicted in table 2.2. For example, there is more than one method involving projected benefits with actuarial liability at the aggregate level. The aggregate entry age normal method and the frozen initial liability method are both examples of this type of method. The distinction between methods at this level is much more subtle than distinctions based on the primary dimensions; also comparing methods based on these names is problematic because in practice, definitions vary across actuaries.

The reporting standards specified by GAO and OMB require that P.L. 95-595 model developers check off from a list of methods the one used in preparing the forecast. As table 2.2 illustrates, the entry age normal and the frozen initial liability methods are the two most frequently reported methods, although there is at least one use of each of the possible types of methods. This summary of method use may be somewhat imprecise given that there is no standard nomenclature in use as discussed earlier.[7]

The choice of a cost method can be partially predetermined by the nature of a plan. For example, accrued benefit methods are usually used in connection with plans that assign calculable parts of the ultimate benefit to number of years of service, or some other measurable incremental factor. This allows for a fairly straightforward calculation of each individual's accrued benefit for a given year. This method is not as appropriate for plans which base their benefit formulae on other criteria, such as the average salary for the final three years of work. For these plans, it is not clear what portion of their benefit an individual has earned in a given year.

---

[7] For simplicity of presentation, we defined and classified methods in this report according to D.M. McGill (1979), a published reference for actuarial professional examinations. A newer pension terminology which also has professional endorsement is available in the report of the Joint Committee on Pension Terminology (1981).

Using different cost methods to calculate the normal cost and unfunded actuarial liability in a given year results in outcomes that depend not only on those methods but on a number of other factors as well. For example, one method may imply higher normal cost for the near term and lower costs later on, while another method may imply the opposite distribution of costs over time. However, plan provisions, the characteristics of plan participants, and the actuarial assumptions all interact with the method used in determining final outcomes. For example, plans that are mature (have been in existence a while) may have already funded a large percentage of the unfunded actuarial liability. For these plans, the difference between with and without actuarial liability methods will not be as significant.

The projection method used in the OASDI cost estimate model differs from the balance sheet method in that projections of factors affecting costs, such as number of covered workers and retirees and the amounts of covered payroll, benefits payable and income are forecasted on a year-by-year basis. The cost of the program is calculated annually as a percentage of covered payroll. Rather than calculating normal cost for the OASDI program, the model calculates average cost as a percentage of taxable payroll over a 75-year period. Because the cost estimates are determined for every year over the 75-year forecast horizon, it is possible to calculate average cost for intervals less than the 75-year total. The actuarial balance of the program is assessed by comparing average cost to its equivalent income measure—the average taxation rate.

Data Sources

The data sources or input for the cost models include historical information on plan participants and future values for model predictors (i.e. the model assumptions). A data set contains information on each participant. For example, information on current employees would include salary history, years of service, age, and rights to benefits. For current beneficiaries, the most important piece of information is the amount of their benefit.

In some cases, the information is aggregated, particularly for the cost models with the largest numbers of participants: models of the OASDI, Civil Service, and Military Retirement Systems. These models aggregate the participants into a number of cohorts (e.g. by age).

All of the cost models use some data external to the model data for the values of their economic and demographic assumptions. Many of the models rely on tables developed by external sources for the various demographic assumptions which include mortality, withdrawal and retirement. Economic assumptions can come from a variety of sources. The P.L. 95-595 reporting requirements currently dictate that all model developers use a 5 percent rate of inflation. The developers of the OASDI model, the CSRS model, and the Military model, all have Boards which have final approval on the economic (and in some cases demographic) assumptions which are used in their forecasts. Some developers do not derive their own economic assumptions, but base them on assumptions used by other cost models of plans which are similar; several developers reported adopting the assumptions of the CSRS or Military Retirement models.

Some developers do not rely on external sources for their assumptions but instead base their assumptions on the experience of the plan participants and characteristics of the plan. The developers for models of the larger plans tend to rely more on their own experience. Further detail on the development of assumptions follows in the next section.

Predictors and Assumptions

Retirement program cost models use an identical set of predictors with a few exceptions.[8] Although the set of predictors is similar, values for those predictors do vary across models. The accuracy of model forecasts depends largely on using economic and demographic assumptions that come as closely as possible to the actual future experience. However, the actuarial assumptions used may not represent an actuary's opinion of the most likely future event for individual assumptions. There are penalties for erring on both sides of the actual result. Optimistic assumptions which make the plan fund appear overly healthy run the risk of having plans which are underfunded and not able to pay future beneficiaries. In addition, optimistic assumptions may lead to a decision to provide benefit increases which may further jeopardize the financial status of the plan.

---

[8] The OASDI cost estimate model uses an open group method and thus has predictors not used by the P.L. 95-595 models. For example, the OASDI model estimates the size of the population into the future and is thus concerned with a population fertility predictor. Since the P.L. 95-595 models are only concerned with current employees and beneficiaries (closed group method), they do not use a fertility rate in their calculations.

Pessimistic assumptions on the other hand run the risk of overfunding a plan and inefficiently using financial resources; overly pessimistic assumptions could lead to cuts in benefits which are not necessary.

The penalty for underfunding is considered greater than the penalty of overfunding and given this, some actuaries select assumptions on the conservative side, or choose a single assumption on which to be conservative. Conservative assumptions or methods are ones that have a negative effect on the financial appearance of a plan and can imply an increase in funding or a reduction in benefits.

Since it is unlikely that actuarial assumptions will be entirely accurate, forecasts typically include an adjustment for the actuarial loss or actuarial gain from previous forecasts.[9] (These terms are defined in the glossary). That is, adjustments are made in the following year's valuation to reflect the difference between actual experience and the assumptions from the previous year. The penalty for inaccurate assumptions is reduced since the assumptions can be changed every year, and there are adjustments to correct inaccuracies in the previous year's assumptions—it is an incremental process. This does not mean, however, that potential errors in some of the core, or central, assumptions will be detected and revised in time to avoid plan funding problems requiring major policy changes. For example, the divergence between forecasted assumptions and actual experience in the 1970s contributed to the social security funding crisis addressed by the 1977 and 1983 Social Security Act amendments. In this instance, the errors in assumptions had major consequences to policymakers.

Assumptions are typically derived through the extrapolation of past experience into the future. An actuary may make some adjustment to a statistical extrapolation based on anticipated changes in the future. Since data on plan experience is often not available, or the plan experience is not very long, some assumptions are derived from standard tables which may or may not be adjusted to reflect specific characteristics of the plan and the population.

---

[9] Standard private pension plan valuation methods such as those used by the P.L. 95-595 models explicitly incorporate this adjustment. There is no such adjustment for the OASDI model, which does not calculate normal cost.

In some cases, assumptions may be designated by the plan sponsor or others. There are separate Boards of Actuaries for the CSRS and Military Retirement System and a Board of Trustees for the OASDI program, which approve assumptions to be used in forecasts. Some plans which are not restricted to assumptions of the Civil Service or Military Boards of Actuaries may adopt one or more of these Boards' assumptions or assumptions recommended by the Social Security Board of Trustees because of similarity between a plan or plan population with one of these larger programs. For example, several of the models for federal employees report using some of the assumptions from these Boards. The Military model used the OASDI mortality improvement assumptions to construct their own unisex mortality tables.

Economic Predictors and Assumptions

The economic assumptions for the P.L. 95-595 models include the inflation rate, the rate of wage increase, and the rate of return on plan investments. Table 2.3 lists the values of economic assumptions used for the 1983 reports for each of these models.

Table 2.3: Economic Assumptions Used in Models of Federal Retirement Programs Reporting Under P.L. 95-595 for the 1983 Plan Year

Retirement system	Rate of of return(%)	Inflation rate[a](%)
1. Civil Service	6.0	5.0
2. Military	6.0	5.0
3. Coast Guard	6.0	5.0
4. Federal Reserve	7.5	5.0
5. Army/Air Force	8.0	5.0
6. Tennessee Valley Authority	7.5	5.0
7. Navy/Coast Guard Resale	8.0	5.0
8. Foreign Service	6.0	5.0
9. Army	7.5	N/A
10. Public Health Service	6.0	5.0
11. Air Force	7.0	5.0
12. Marines	7.5	5.0
13. Louisville, KY FCB	6.0/7.0[b]	N/A
14. St. Paul, MN FCB	6.5	5.0
15. Omaha, NE FCB	7.5	5.0
16. Columbia, SC FCB	7.0	N/A
17. St. Louis, MO FCB	7.5	5.0[c]
18. Wichita, KN FCB	8.0	5.0[d]
19. Sacramento, CA FCB	8.0	N/A
20. Spokane, WA FCB	8.0	N/A
21. Judiciary	7.0	5.0
22. Baltimore, MD FCB	9.0	N/A
23. Austin, TX FCB	6.5	N/A
24. Springfield, MO FCB	7.0	N/A
25. Jackson, MI FCB	6.0/13.66[e]	N/A
26. Jackson, MI FCB, Production Credit Association	6.0/13.69[e]	N/A
27. National Oceanic and Atmospheric Administration	6.0	5.0
28. Federal Home Loan Mortgage	8.0	None reported
29. Norfolk Naval Shipyard	8.0	5.0
30. Navy Morale, Welfare and Recreation	8.0	5.0
31. Tax Court	7.0	5.0

[a] The inflation rate assumption is not applicable for those plans which do not have indexation of benefits to inflation.

[b] The lower rate is used to calculate annual cost; the higher rate, accumulated plan benefits.

[c] A lower rate (3 percent) is applied to benefits for those who retired prior to 5-1-74.

[d] The actual assumption is that the rate will be greater than 3.

[e] The higher rate is applied to calculations concerning those who retired prior to 1-1-84; the lower, to all other calculations.

There is no variation in the inflation rate assumption for those models using one because the GAO-OMB reporting requirements mandate that plans use a five percent inflation assumption. Some model developers do not use inflation rate assumptions because it is only important for plans whose benefits are indexed to the consumer price index.

The inflation rate is not allowed to vary for the P.L. 95-595 reports, but for valuations where it is allowed to vary, a higher inflation rate (holding all other assumptions constant) is a more conservative assumption—it would tend to make the normal cost for the current year higher.

The rate of return on plan investments was reported in all of the model documentations. As table 2.3 shows, estimates ranged from six percent to nine percent. Plans invest their assets differently; valuations were made at different times of the year; and some model developers might deliberately make their forecasts more or less conservative.

The rate of return estimates how much the fund will earn. It is also the rate at which future benefits are discounted in the calculation of normal cost. Changing this assumption has the most impact for plans that build up funds of some size. A lower rate of return (holding all other assumptions constant) is a more conservative assumption—it would tend to increase the normal costs.

A third important economic assumption in these models is the assumed rate of wage increase. Wage increases are composed of three parts: increases due to changes in cost of living, general productivity and promotion and other merit awards. Thus the wage increase assumption may not always be reflected in a single number. It can change over time and be different for different demographic groups of employees. A higher rate of wage increase assumption (holding all other assumptions constant) is a more conservative assumption—it would tend to increase normal cost. Some information on wage increase assumptions was reported in model documentation but figures across models are not comparable, as models use various combinations of the wage increase components. It is not entirely clear in the model documentation which component is being reported.

Unlike the P.L. 95-595 models, projections from the OASDI cost estimate model use four well defined sets of assumptions about future economic and demographic trends. These are known as Alternatives I, II-A, II-B, and III. The Alternative I assumptions reflect an optimistic view of the factors that determine OASDI costs. Forecasts using these assumptions indicate lower costs and better financial status for the program. The Alternative III assumptions reflect a pessimistic view of the same factors and this view is reflected in forecasts of higher costs and poorer financial condition. Alternatives II-A and II-B reflect intermediate levels of the assumptions, with II-A values representing a future economy much like that experienced in periods of robust economic growth and II-B values less optimistic than that. The Alternative II-B assumptions are the ones recommended by the model developers as the best set for evaluating the financial status of the OASDI program. These assumptions, published along with the annual OASDI forecasts, are widely applied by other modelers to achieve consistency where possible with the OASDI model.

Table 2.4 presents some of the actual II-B assumptions for the 1983 OASDI forecast. As table 2.4 indicates, the assumptions are dynamic, changing over time. For intermediate years, the rates are generally derived from smooth trends from the short-term to the ultimate rates. The exact procedure used to establish these trend lines varies across assumptions. Some trends are estimated by expert judgment, others by various statistical curve fitting procedures.

The dynamic nature of the assumptions, the fact that some assumptions, such as the wage rate increase, apply only to OASDI covered employment, and the different effects of assumptions on the unique outcomes forecasted by the OASDI model make it difficult to compare these assumptions with those used in the models of federal employee retirement programs. For example, a glance down column 3 of table 2.4 shows that OASDI's "most likely" rate of inflation (indicated by the consumer price index) only exceeds the five percent rate mandated for use in federal retirement forecasts for the 1985 year. If all other assumptions were equal and forecast objectives were equivalent, this difference would produce higher normal cost estimates for the federal retirement programs relative to the OASDI program. However, the OASDI model does not evaluate normal cost. Its primary outcome is the long range average cost rate. In general, higher inflation rates make the cost rate lower. Thus, for the OASDI model, lower inflation rates are more conservative rather than less conservative.

Table 2.4: Alternative II-B Economic Assumptions Used in the 1983 OASDI Forecast

	Assumptions		
Calender year in which the rate takes effect	Wage rate increase[a]	Rate of increase in consumer price index	Average unemployment rate
1983	4.6%	3.1%	10.1%
1984	4.6	4.4	9.1
1985	5.5	5.3	8.3
1990	5.6	4.0	6.5
2000 (and later)	5.5	4.0	5.5

[a] The assumed wage rate increase is for OASDI covered employment.

Source: 1983 Annual Report of the Board of Trustees, p. 37.

In general, however the effects of economic assumptions on a forecast are determined in many cases not by any individual assumption, but by the relative differences among assumptions. Some developers select assumptions to maintain expected differentials among rates. For example, given the constraint of using a five percent inflation rate, the Military and CSRS developers used differentials to select values for other economic variables. Interpreting the appropriateness of the economic assumptions therefore may involve an examination of the differences between the inflation rate, the rate of return, wage increase and other assumptions as well as their actual values.

Demographic Predictors and Assumptions

Demographic assumptions determine what happens to a participant population over the course of the forecast period. The most common assumptions involve when participants will die, retire, or leave their job before retirement. There are other assumptions as well. Plans with disability benefits have assumptions about rates of disability, and plans with survivor benefits have assumptions concerning rates of mortality for survivors. There are other types of assumptions depending on the characteristics of individual plans. For example, future fertility rates in the general population are an important demographic assumption for models using an open group valuation method as in the OASDI cost estimate model. These rates affect the estimated numbers of future workers with OASDI coverage and the estimates of the number of beneficiaries with dependents benefits. As opposed to the economic assumptions where one number typically reflects the assumption, demographic assumptions are specified differently for different segments of the population (e.g. different age-sex groups).

Most models use published tables as the source for their mortality assumptions. These mortality tables (or, as they are sometimes called, life tables) give the probability of death for different ages. Mortality tables differ according to the populations they cover. There are tables based on the entire U.S. population developed by the Bureau of the Census, tables based on the U.S. population which are specifically derived for use in the OASDI cost estimate model and tables based on the experience of individuals covered by annuities issued by life insurance companies. Tabled assumptions that are not based on plan experience may be adjusted to reflect characteristics of the plan population. These adjustments are described in each reported forecast.

The majority (19 out of 31) of models of federal retirement programs used the 1971 Group Annuity Mortality Table which is based on the experience of individuals receiving life insurance annuities. Two models rely on an earlier version (1951) of this table; one, a more recent version (1983); and two, a similar one, the 1971 Towers, Perrin, Forster and Crosby tables. The Military Retirement model used unisex mortality tables developed from plan experience and OASDI II-B assumptions regarding rates of improvement in mortality. Of the remaining models, two derived their mortality rates from the experience of their participants over a designated time period, two on the experience of officers in the Military Retirement System, and two on 1984 mortality tables.

OASDI mortality rates are developed as a separate stage in the modeling process, along with other OASDI population assumptions. The resulting tables are published and thus are available for use by other modelers.

Many of the models continue to use old mortality tables. This is not an uncommon practice in the pension community because mortality rates change slowly and use of more recent tables may not have a significant impact on the final forecast for specific plans. However, mortality rates have been declining in general and applying lower (or more recent) rates of mortality (holding all other assumptions constant) is a more conservative assumption as it implies higher normal cost for the valuation year.

Plan sponsors are not currently required to report retirement assumptions although some do. For the three largest programs—OASDI, CSRS, and Military Retirement—these assumptions are developed from program experience using various trend extrapolation procedures. Developers of the OASDI and CSRS models reported to us that these assumptions are modified when the models are used to examine the effects of proposed program changes which would alter the relationship between the amount of benefits and the age of retirement. Changes of this type could affect the future retirement decisions of workers and subsequently the costs of the programs. For both models, the impact of policy change on retirement rates is determined by expert judgment. Some of the retirement decision models described in chapter 3 were developed specifically to estimate these impacts. We are unaware of any studies that compare the estimated impacts with the retirement impact assumptions used in the OASDI and CSRS models.

Seven of the federal retirement models cited "plan experience" as the source of employee withdrawal assumptions. The remaining plans report using either standard tables (developed by an actuarial firm) or do not report the source of their assumptions. For the OASDI program the withdrawal assumption is not directly relevant. Instead the model estimates the number of covered workers, using procedures described in appendix II of the supplementary volume of this report.

Other Predictors

Some assumptions may not fall neatly into the category of demographic and economic assumptions. For example, numerous assumptions about future labor force participation rates for OASDI covered employment, work patterns, salaries and male-female wage differentials, are used in the OASDI model to estimate what benefits will be payable to future retirees. These assumptions are developed using a variety of methods from expert judgment to statistical simulation. They are described in more detail in the supplementary volume of this report.

Analytic Dimensions

Next, we review the models of the 32 federal retirement programs in terms of the three analytic dimensions defined in chapter 1: documentation (availability of user oriented documentation), maintenance (frequency of model updating and revision), and operational validity (procedures used by model developers to monitor the divergence between real world and model outcomes).

Documentation

In this review, we focused on publicly available documents which describe each model and how the model is used to produce forecasts. Thus, our summaries of model documentation refer to the documentation that a potential model or forecast user might examine.

For the three largest models of retirement program costs (the OASDI, CSRS, and Military Retirement models), documentation consisted of in-house publications and reports of annual valuations or forecasts. For the smaller models of federal retirement programs, the only documentation source we examined was the annual report mandated by P.L. 95-595. The reporting requirements under P.L. 95-595 ensure some consistency in report contents across models. The focus of the report is the forecast itself. However, modelers are required to indicate the actuarial methods and the assumptions used to produce the forecast.

We found problems in interpreting the information on methods and assumptions reported by P.L. 95-595 model developers because a standard nomenclature for actuarial methods does not exist and definitions of assumptions across models can also vary. An example of the latter problem is the wage rate increase assumption. A value for this assumption could be based solely on expected general schedule pay rates or could include expectations concerning merit pay and promotions. Although some developers supplemented their reports to clarify information on assumptions and methods, this was not done consistently.

For the three larger cost estimate models, we found variation in the amount and completeness of documentation. Documentation for the Military Retirement Model was complete, including information on the past accuracy of demographic assumptions and descriptions of model revisions made to correct for those inaccuracies.

Not unexpectedly, documentation in terms of numbers of publications, was largest for the OASDI cost estimate model. However, the model documentation was not complete. An important sub-model (the short-run cost estimate model) was not documented at all and we found no single source of information documenting the procedures used for the entire model. Revised versions of the OASDI cost estimate model, used to assess proposals which led to the enactment of the 1983 Social Security Act Amendments, were also not documented at all. While model developers were planning to document the short-run estimation procedure, there were no plans for documenting the revised versions of the model and developers were unsure about how these versions would be used in the future.

The amount and completeness of documentation for the CSRS model was even less than that of the other large plans. A 1982 GAO review concluded that documentation for the CSRS model was not adequate.[10] Developers reported to us that they have taken some steps since that review to supplement the amount of in-house documentation, but this documentation is not published. These improvements were being made in documentation sources, such as the computer code, which we did not evaluate.

Maintenance

Cost models are updated annually to reflect changes in assumptions and in law, with models of the largest programs undergoing more substantial revision than other models. In some instances, revisions are based on changes in population characteristics. However, other revisions are based on methodological changes. For example, the CSRS model was modified to produce both static (no wage inflation) and dynamic cost estimates. The Military model was revised to capture more correctly the varieties of possible entitlements to the program; the OASDI model was changed to include new methods for estimating both future benefits and future revenues.

The extent of model revision can be estimated roughly by knowing whether the actuary performed a full or partial valuation. (Valuation as used here refers to the process of producing a forecast rather than the forecast itself.) A full valuation involves an assessment and adjustment (if necessary) of all assumptions and methods used in valuation. A partial valuation focuses only on some of the assumptions and methods. There is no standard defining a partial valuation and thus the extent of new assessment of methods and assumptions can vary widely for those doing a partial valuation. It is common practice for actuaries to make a full valuation every three years, or some other specified interval, with partial valuations done in the intervening years. The P.L. 95-595 sponsors do not report on the extent of valuation, and it would be difficult to do so given the lack of a disciplinary standard. It is known that for one model, the CSRS model, full valuations are prepared every five years when the major plan report is produced.

In addition, the CSRS and OASDI models undergo temporary revisions to evaluate the effects of proposed reforms for congressional and executive agency personnel. Many such changes to the OASDI model were needed to evaluate proposals by the National Commission on Social Security Reform.

---

[10] See Inadequate Internal Controls Affect Quality and Reliability of the Civil Service Retirement System's Annual Report, AFMD-83-3. Washington, D.C., October 22, 1982.

Validity

There are three ways to examine the potential sources of forecast error for the cost models and each has limitations. The first is to test the historical accuracy of forecasted outcomes. The second is to observe the historical accuracy of model assumptions, and the third method is to examine the sensitivity of model results to changes in assumptions.[11] One major weakness of the first two methods is that they often do not provide fair tests of the accuracy of a given model.[12] The third method provides only indirect information on potential forecast error.

As discussed earlier, the outcomes typically reported by cost models are not "forecast" but determined. The determination of these outcomes—the normal cost, the present value of future benefits, and the actuarial liability—involves generating forecasts of benefits and revenues for each year into the future. The historical accuracy of these forecasts could be tracked. The forecast objective is to estimate funding needs over the life of the plan and thus the ideal assessment would be to examine the accuracy of the entire forecast at the end of the forecast horizon. However, for the OASDI model, the forecast horizon is 75 years and for others it may be as long as 50 or 60 years.[13] After that amount of time, there may be little interest in such an accuracy study. Tracking the accuracy of intermediate forecasts is difficult because most models do not provide annual output. To the extent such outputs are possible from the models or have been recorded over time, the historical accuracy of forecasts could be assessed. In our search, we identified only one analysis of the historical accuracy of a cost estimate model. That analysis was for the model used in 1935 to estimate future costs of the OASDI program (Myers, 1983).

---

[11] The sensitivity of a model is not always an undesirable property. Circumstances under which sensitivity is desirable are situation specific.

[12] It is generally agreed that the accuracy of a model's forecasts can only be fairly measured by a large number of forecasts over a relatively long time period. (Success or failure in one or two forecasts can be attributed to chance, assuming some random component in the forecast error.) Given the dynamic nature of the modeling process, it is difficult to refer to the accuracy of the OASDI model, for example, because the models and individuals in charge of them are changing over time. Evaluators tend to deal with the problem by referring to the forecast accuracy of models associated with a particular developer or sponsor.

[13] The forecast horizon for the P.L. 95-595 models is the time up to which the last employee (employed as of the valuation date) receiving benefits dies. If there are current employees as young as 25, the forecast horizon could be as long as 50 or 60 years.

For the OASDI cost estimate model, sufficient information is published annually to evaluate shorter-range forecasting accuracy. Although much has been written about potential errors in these forecasts, we could not find any evaluations of them. We also believe no analyses exist of the forecasting accuracy of the remaining 31 models.

A second method for examining potential sources of forecast error is to assess the accuracy of model assumptions. The assumptions which are forecasted for future years are key determinants of the outcomes. Model developers track the error in their assumptions when they calculate actuarial gain and loss. The size of the actuarial gain and loss can be used to some extent to test the short-range predictive ability of a model, although large actuarial gains or losses could result from factors outside the model such as changes in the rules of the pension plan which would need to be identified. This information is not reported in the P.L. 95-595 reports.

Assessing the accuracy of model assumptions is a problem for models which use static assumptions; that is, assumptions which remain constant (or change very little) over the years of the valuation. The goal is to approximate an "average" value for the assumptions over the forecast horizon. Dynamic assumptions, on the other hand, could change significantly over time as they try to capture year to year variation. The accuracy of static assumptions could only fairly be assessed at the end of the forecast horizon in order to see how reasonable a particular assumption was, on average. We did not find any studies of assumption accuracy for the models used in preparing the P.L. 95-595 reports, which primarily use static assumptions.

Since the OASDI model uses dynamic assumptions, accuracy of assumptions could be tested, although even changes in these assumptions are forecasted as gradual trends. As part of a 1983 review of the integrity of the forecasts made for the OASDI program during the period 1973-82, we examined the accuracy of the model assumptions for the first nine years in the forecast horizon. A partial summary of the results of our review is provided in table 2.5. We concluded that during that period, the actual experience for unemployment and CPI was higher than had been forecasted causing actuarial projections to understate costs and overstate revenues. We provide a full discussion of the inherent difficulties in accurately projecting economic and demographic conditions in our 1986 report on Social Security projections.[14]

Table 2.5: Average Difference Between Actual and Forecast Values for Selected Economic Assumptions for OASDI Trustee Report Years 1973-81

		Economic assumptions		
Year in forecast horizon	Number of observations[a]	Increase in consumer price index	Increase in wages	Rate of unemployment
First	9	0.6%	−0.3%	−0.1%
Second	8	3.1	−0.4	0.5
Third	7	3.8	−0.5	1.4
Fourth	6	4.6	0.8	1.6
Fifth	5	5.9	1.8	1.7
Sixth	4	7.1	2.5	1.9
Seventh	3	8.4	3.3	2.2
Eighth	2	9.0	3.5	2.6
Ninth	1	7.6	3.6	3.1

[a] The analysis of accuracy was done in 1982, so there were nine first year forecasts that could be examined (one for each of the report years 1973-81). There are only eight observations for the second year forecasts because the actual value for 1983 was not available to compare with the forecast value in the 1982 report.

Source: U.S. General Accounting Office. Social Security Actuarial Projections, HRD-83-92. Washington, D.C., September 30, 1983, pp. 7-14.

A third method of assessing potential forecast error is to conduct sensitivity analysis. Such an analysis for cost models would involve manipulating assumptions, one at a time, to determine the effect on model outcomes. A sensitivity analysis can help provide confidence bands around model results, and can be particularly useful in light of information on past error in and variation of particular assumptions. Bartlett and Applebaum (1982), who examined errors in the 1970-79 II-B assumptions for the OASDI model, concluded that errors in economic assumptions as large as those of the early 1970s can produce five-year cost estimates that differ from actual experience by as much as 40 percent of annual benefit payments.

In general, we did not find results of sensitivity analyses of model assumptions (the P.L. 95-595 reporting requirements do not request such results), although some of the model developers suggested that they do conduct such analyses. Additional documentation for the Military Retirement model contained some information on sensitivity analysis. In contrast, results of sensitivity analyses are routinely reported for the OASDI model where four sets of assumptions are used to generate four forecasts with outputs ranging from optimistic to pessimistic. However, just presenting optimistic and pessimistic assumptions does not guarantee that results will fall within that range, and if the spread between the sets of assumptions is too great, they may not may not be useful.[15] In spite of limitations, testing the sensitivity of results and using a range of forecasted values rather than a point estimate may be better ways to present and use forecast results.

---

[14] U.S. General Accounting Office. Social Security: Past Projections and Future Financing Concerns, HRD-86-22. Washington, D.C., March 11, 1986.

Summary

In this chapter, we reviewed 32 cost models of federal retirement programs: 31 whose sponsors report annually under P.L. 95-595 along with the model of the OASDI program. The primary objective of the models' forecasts is to ensure that the programs are soundly funded for the future. It is generally agreed that the penalty for underfunding a plan—not being able to pay future benefits—is greater than the penalty for overfunding it—unnecessary benefit reduction—and thus modelers prefer procedures that minimize forecast error in the direction of underfunding. Plan provisions, the characteristics of plan participants, the actuarial methods and assumptions (predictor values) all interact in the determination of the model final outcomes—normal cost and the actuarial liability for the P.L. 95-595 models, and average cost and the trust fund balance for the OASDI model. While the first two factors are fixed for a given model, the developer is free to select a method and select or estimate the assumptions. Conservative or pessimistic assumptions or a combination of method and assumptions that yield conservative forecasts provide lower risks of underfunding.

For P.L. 95-595 models, the inflation rate assumption is controlled by GAO-OMB requirements and for the 1983 plan year was 5 percent, higher than that used in the OASDI model. The rate of return varied from a low of 6 percent for a number of plans to a high of 9 percent for the Baltimore Farm Credit Bank plan. We did not compare wage increase assumptions across models because developers did not always report which components (cost of living, productivity and merit increases) of this assumption were included in their estimated rate. Two-thirds of the models used externally developed mortality tables published prior to

---

[15] Light (1983) noted that the range between the optimistic and pessimistic assumptions for the OASDI model has been increasing over time in response to previous error in assumptions.

1972 for mortality assumptions and only four derived plan-specific mortality rates.

The amount of model documentation varied across models. The P.L. 95-595 reporting requirements dictate the minimum extent of documentation for 31 of the models. Some developers supplemented the basic required information in their reports. We examined additional documentation which was published for the three largest models and found substantial variation. Documentation for the Military Retirement model was complete and CSRS and OASDI, incomplete. Documentation was often difficult to interpret because there is no nomenclature for actuarial methods in standard use and all assumptions are not operationally defined for ease of comparison. In addition, there was little published detail on methods for the small plans and for CSRS and no single collective documentation source for the OASDI model methods.

All of the models are maintained relatively frequently because they produce annual forecasts, although the extent of revision in any given year varies across models. Full valuations which include comprehensive updating and revision are done cyclically but not annually for most models.

Information on the potential for forecast error in these models is seriously lacking. We found only one evaluation of the long-term historical accuracy of a model and none of short-term forecast accuracy. Actuarial gains and losses attributed to changes in assumptions are not routinely reported under P.L. 95-595 and we did not find any studies of assumption accuracy for these models. For the OASDI model, we and others recently examined the accuracy of model assumptions and found them overly optimistic in the 1970s.[16] Estimates of forecast error were also not provided for these models although OASDI and Military Retirement model developers published sensitivity analyses.

For the largest programs (OASDI, Military, and Civil Service) sufficient statistics are available to track forecast accuracy, although it has not been done. It may not be possible to do so for the smaller programs. Given the lack of information on forecast accuracy, sensitivity analyses which provide a range of estimates (rather than a point estimate) may provide information on potential forecast error.

---

[16] See GAO (1986), GAO (1983b), Light (1983) and Bartlett and Applebaum (1982).

Chapter 3

Models of Retirement Decision Behavior

Models of retirement decision behavior can provide information about when and why people retire. Retirement can be defined in many ways, the least stringent of which is accepting pension benefits and the most stringent, withdrawing totally from the labor force. Models of these decisions can be used to predict future behavioral trends under existing retirement policy or under alternative policies. Results also are useful information sources for projections of retirement income and retirement program costs.

In this chapter we review 35 empirically estimated models of the retirement decision. These models are individually described in the supplementary volume of this report. Most of these models were developed to estimate the relationship between the availability and amount of social security benefits and the retirement decisions of workers. Many of these models can produce estimates of what changes in retirement decisions would be expected if benefits were changed, and they can predict the effects on retirement of changes in worker characteristics, such as health.

Unlike the cost estimate models reviewed in chapter 2, which were developed specifically for federal government use, these models were developed by private researchers in the academic community. As we mentioned in chapter 1, this chapter includes models of both public and private sector civilian employees. Although these models share the common objective of depicting the retirement decision making process, they differ in their approach to achieving that objective. The models vary in outcomes which are predicted, methods of estimation and model structure, data sources, and selection of predictors.

Table 3.1 lists these models. Since they do not have names, we refer to them by the name of the model developers. In those cases where the same model developers have more than one model, the date of initial model publication is used to identify separate models. Models are listed and numbered chronologically by publication year and alphabetically within a year. In the remainder of this chapter we provide some examples of how these models have been used for retirement policy analysis, describe the models along four dimensions (outcomes, methods, data sources and predictors) and provide information on the availability of model documentation, on model maintenance and on how developers treated questions concerning the models' operational validity. The chapter concludes with a summary of the implications of these descriptive and analytic dimensions for model use.

Table 3.1: Models of Retirement Behavior

No.	Date	Name
1	1976	Burkhauser ba-lfp model of auto workers
2	1977	Boskin lfp[a] model
3	1977	Quinn lfp model
4	1978	Boskin-Hurd ba[b]-lfp model
5	1978	Pellechio lfp model
6	1979	Schmitt-McCune ba-lfp model of Michigan civil servants
7	1980	Barker-Clark lfp model
8	1980	Burkhauser ba model
9	1980	Burtless-Hausman ba-lfp model of federal civil servants
10	1980	Clark et al. joint lfp model
11	1980	Gordon-Blinder lfp model
12	1980	Henretta-O'Rand lfp model of women
13	1981	Burkhauser-Quinn lfp model
14	1981	Gustman-Steinmeier model
15	1981	Hurd-Boskin lfp model
16	1982	Gustafson lfp-ba
17	1982	Hamermesh lfp model
18	1982	O'Rand - Henretta age of retirement model
19	1982	Slade lfp model
20	1983	Anderson-Burkhauser lfp health model
21	1983	Fields-Mitchell age of lfp model
22	1983	Gustman - Steinmeier model
23	1983	Honig-Hanoch lfp model
24	1983	Mitchell-Fields ba model
25	1984	Anderson et al. retirement plans model
26	1984	Burtless lfp model
27	1984	Burtless-Moffitt lfp model
28	1984	Diamond-Hausman hazard model
29	1984	Diamond-Hausman probit lfp model of the unemployed
30	1984	Diamond-Hausman competing risks lfp model of the unemployed
31	1984	Gohmann-Clark age of ba model
32	1984	Gohmann-Clark lfp model
33	1984	Hausman-Wise Brownian motion lfp model
34	1984	Hausman-Wise hazard model
35	1984	Kutner age of ba-lfp model of California educators

[a] labor force participation (lfp)

[b] benefit acceptance (ba)

Background and Use

Approximately one-third of the 35 models have been applied in policy experiments, including backcasts of the effects of 1969 and 1972 social security benefit increases on retirement behavior (predictions of events that have already occurred) and forecasts of the effects of various potential private pension and social security policy changes.

Proposed social security policy changes which have been assessed include changing the normal age of retirement, changing the retirement age incentive structure of benefits, eliminating or revising earnings test policies, delaying the cost of living adjustment, and assessing the overall and individual effects of the 1983 Social Security Act Amendments.

There have been fewer policy experiments on issues other than social security. Two models estimated the effects of the 1978 legislated change in allowable mandatory retirement rules (Age Discrimination in Employment Act Amendments). One model forecasted the short-term effects of changing the age requirements for receipt of a federal pension and eliminating "windfall" benefits to federal employees who are also covered by the OASDI program. Another model forecasted the effects of benefit formula changes in a state administered pension plan.

In addition to these policy experiments, three models have been used for other types of forecasts. One of these has been part of the DYNASIM model (reviewed in chapter 4) since 1981; thus DYNASIM forecasts are based in part on its results. Another model forecasted retirement patterns under separate assumptions of long term economic growth in real wages and the elimination of private pension income. The effects of onset of a long-term health problem at age 55 were also estimated with this model. The third model examined the effects on long-run OASI cost estimates of using a behavioral response model in place of actuarial retirement assumptions.

Descriptive Dimensions

Models of the retirement decision differ from the models of retirement program costs reviewed in chapter 2 on outcome variables and in the amount of estimation underlying both types of models. Models of the retirement decision involve more variety of estimation than cost models, because both the factors and the manner in which they influence behavior are free to vary for these models in contrast to the cost models. Thus major components of the modeling process involve specifying the factors and specifying the way in which they influence behavior. Another important component of the process is testing the model specification. This involves application of statistical estimation methods to real samples of individuals, whose retirement behavior and characteristics on selected factors have been observed and recorded.

Each phase of the modeling process affects interpretation of modeling results and the models vary in how each phase is accomplished. Thus, our descriptive review summarizes the class of models on all four of the descriptive dimensions described in chapter 1: the specific outcome variables forecasted by the model, the method of estimation, the data sources or samples on which the models have been tested, and the factors included in the models as predictors of behavior.

Outcomes

The most popular definition of "retirement" (80 percent of the models) is related to an individual's labor force participation (or lfp, see again table 3.1). Measures include complete withdrawal from the labor force, partial withdrawal from the labor force, a discontinuous drop in hours worked below some specified limit and quitting the main job.

Other definitions of retirement are self-assessed retirement status (five models) and receipt of pension or retirement income (nine models). Of the latter models, three defined retirement status by receipt of social security benefits; one, a federal pension; two, a state pension; two, a private pension, and one used receipt of either a private pension or social security.

Although there is some correlation between receipt of pension income and labor force participation, the relationship is not perfect. This is explicitly recognized in the Social Security program by the earnings test, which allows workers to receive benefits and work so long as earnings do not exceed a specified limit. Typically, with other pensions, workers must leave their main full-time job in order to receive pension benefits but they are not prevented from accepting alternative employment, and many workers do.

Some developers have modeled alternative definitions of retirement. For example, Burkhauser (no. 1) predicted both early private pension benefit acceptance and labor force participation. Likewise, Gohmann and Clark (no. 31) examined age of social security benefit acceptance and years to labor force withdrawal after benefit acceptance. Other examples include Honig and Hanoch (no. 23), who predicted labor force participation status, reduction in work effort and partial retirement (part-time work); Diamond and Hausman (no. 28), labor force participation status and self-assessed retirement status; Burtless and Moffitt (no. 27), age of retirement and post-retirement work effort; and Gustafson (no. 16), five definitions of retirement, including benefit receipt, earnings under the earnings test limit, labor force participation, working a half year or less and leaving the main job.

Others have handled the multiple definition problem by defining retirement as simultaneously meeting more than one definition. For example, Boskin and Hurd (no. 4) classified individuals as working with no retirement income, receiving social security benefits and working, or not working and then modeled the probability of being in one of the three categories. Burtless and Hausman (no. 9) similarly defined retirement for federal workers as accepting a pension and withdrawing from the labor force or accepting a pension and taking a job in the private sector. Schmitt-McCune (no. 6) defined retirement as accepting pension benefits and leaving the main job.

A few of the models depict retirement as a time for multiple decisionmaking. In these models, the decision to retire is modeled simultaneously or jointly with other decisions. Examples of joint decision models are the Clark et al. model (no. 10) of the joint decisions of husbands and wives to withdraw or participate in the labor force and the Hamermesh model (no. 17) of joint work reduction and consumption decisions.

If interest is primarily in understanding the effects of retirement on future labor supply, models predicting labor force participation are more appropriate. On the other hand, if interest is primarily in the effects of retirement on the costs of retirement income programs, then models using benefit acceptance as the outcome variable are more appropriate. Although the best measures of each of these outcomes are direct ones, some developers have used an indirect measure, such as self-assessed retirement status reported by surveyed individuals. Diamond and Hausman (no. 28) and Gustman and Steinmeier (no. 14) both reported that this less direct measure of retirement behavior gave comparable results to ones obtained using more direct measures.

Methods

The majority of models (32 of 35) were developed from the perspective of economic life cycle theory.[1] The life cycle model is a general model of human decisionmaking based on economic theory. The general model assumes that life choices are based on attempts to maximize the utility (satisfaction) realized from lifetime consumption and leisure, given the opportunities available to an individual. With respect to retirement, the model assumes that workers select a retirement age that maximizes the utility from consumption and leisure for their remaining years. The model is applied by trading off available income from different sources (e.g. wages for continued working, available retirement income from pensions, and asset accumulations) against one another and against the utility from leisure, which may vary with age. Many models include demographic characteristics of workers, such as age, race, marital status and education, to capture some of the individual differences in preference for retirement or leisure.

---

[1] See Modigliani and Brumberg (1955).

Structural life cycle models of retirement use estimation procedures that are linked closely to the mathematics of life cycle theory. The structural models yield equations in which theoretical constructs (such as parameters of a utility function representing the preferences of the individual) are related to retirement outcomes. The values of these theoretical constructs are estimated statistically from information on the alternative courses which were available to the individual and the course which was actually chosen. That is, parameters representing individual preferences are estimated as being the most likely values that are consistent with the opportunities the individual faced and the behavior which was observed. Once the parameters of this preference function are estimated, the estimated values can be used to simulate (or predict) how individuals would respond to changes in the rules of retirement programs or other opportunities they faced. Six of the models we review are structural life-cycle models of the retirement decision (models 11, 21, 22, 24, 26 and 27).

Reduced form models estimate the statistical relation between certain predictors and the retirement decision which is observed. The relationships estimated have not had a one-to-one correspondence with the mathematical specification implied by life cycle theory. That is, actual utility function parameters are not estimated. Since theoretically it is these parameters which affect retirement decision-making, a reduced form model that does not estimate them directly could have high current explanatory power but low power to predict the consequences of policy change on decision-making in the future. Twenty-six of the 32 life cycle models we review are reduced form models.

Four of the reduced form models are longitudinal models—they focus on the transitions in work/retirement behavior over several years. These models apply mathematical distributions known to specify the behavior of certain physical objects to the retirement behavior of individuals. These methods are described more fully in the technical descriptions of models nos. 2, 8, 33 and 34 in the supplementary volume of this report.

The remaining three models were developed from other theoretical perspectives (model 6, psychology; models 12 and 18, sociology) but were estimated with statistical techniques similar to the majority of the reduced form life-cycle models.

Data Sources

Each of the models of the retirement decision was tested by the developers on one or more samples of individuals. In each case, the sample data were collected prior to specification of the model.[2] Thus, when comparing the models on other factors, such as choice of outcome variable or selection of predictors, it is important to remember any differences in data sources which might constrain the model's specification. A given model's validity or ability to explain observed behavior might increase substantially if it were retested on more suitable data.

Table 3.2 summarizes the data sources for each retirement decision model. As table 3.2 shows, the Longitudinal Retirement History Survey (RHS), sponsored by the Social Security Administration in 1969-79, is by far the most frequent data source: 66 percent of the models drew on some data from this survey. Four models used the National Longitudinal Surveys of Labor Market Experience (NLS); one, the Michigan Panel Study of Income Dynamics (PSID); and two, data from the Current Population Survey (CPS).

---

[2] An alternative approach would be to specify the model and then collect data that precisely meets the model's requirements. The latter approach provides the best test of the model's ability to predict behavior. However, data collection costs are sufficiently high that model developers have accepted the constraints that occur when using pre-collected data. For example, few existing surveys contain the kind of detailed information on individual private pension coverage that modelers would like and thus the effects of pensions on retirement behavior are estimated more approximately than desired.

Table 3.2: Data Sources for Retirement Behavior Models

Source	Total	Model Nos.[a]
1969-79[b] Longitudinal Retirement History Survey (RHS)	23	(Nos. 3,4,7,10,11,12,13,14,15,17, 18,19,20,21,22,23,25,26,27,31, 32,33,34)
1966-78[b] National Longitudinal Surveys of Labor Market Experience (NLS)	4	(Nos. 16,28,29,30)
1973 Current Population Survey (CPS)	2	(Nos. 5,8)
1968-72 Michigan Panel Study of Income Dynamics (PSID)	1	(No. 2)
1965-67 Barfield-Morgan United Auto Worker Surveys	1	(No. 1)
(1979[c]) Michigan Civil Service Surveys	1	(No. 6)
1976 Administrative Data File on Federal Workers	1	(No. 9)
1972, 1977 Terman Study	1	(No. 17)
1980 California State Teachers Retirement System Survey (STRS)	1	(No. 35)
1978 Department of Labor (DOL) Benefit Amounts Survey	1	(No. 20)

[a] Model identification numbers are in parentheses. Model No. 17 used two sources.

[b] Not all models used information from all survey years. Refer to the supplementary volume of this report for precise dates used.

[c] Initial date of model publication based on this source.

In general, developers did not provide detailed information on data quality. They frequently reported, for example, that missing data were imputed but provided no information on the percentage of cases with missing data. We see this as an important omission because of the variable quality of both extant and newly collected survey data. We reported on data quality problems in the RHS, noting the high frequency of missing data, incredible values on many income variables, incorrect industrial codes in the RHS manual and numerous other problems.[3] Thus, substantial and costly efforts may be required to diagnose and prepare large sample survey data for reliable use in modeling, and summaries of these procedures are needed to evaluate model outcomes.

As table 3.2 indicates, all but one model were developed on data collected prior to 1980. The single exception is the Kutner model based in part on a 1980 survey of California educators. The use of dated information even in the most recently developed (1984) models is due in part to the continued popularity of the RHS which discontinued data collection in 1979, and in part to the time it takes for collected data to become available in a useful form for modeling.

---

[3] Data from the Retirement History Survey, GAO/IPE-82-5. Washington, D.C., July 6, 1982.

Caution is needed in generalizing results from these models, based on dated information, to future populations. Doing so requires the assumption that the same factors will continue to affect retirement in the same way in the future. The RHS respondents were all approaching retirement age in 1969 (ages 58-63). The last decade and a half has seen much social change. The economy is less predictable than it was prior to the 1970s; the social security program itself has undergone several revisions; and work patterns, especially among women, have also been changing. Continued change can be expected for the future. It is not clear how much these factors have or will influence the retirement decisions of present or future workers.

The majority of models were estimated on fairly large numbers of observations. Twenty-four of the models used over 1000 observations and thirty, over 500. Two did not report estimation sample size.

For most of the models, additional selection procedures were used to develop the sample for the model estimation. These procedures consist largely of partitioning the individuals in characteristics, such as the sex of respondent, and then testing the model on one or more of the partitioned groups. One rationale for doing this is to simplify the model specification by reducing the variability among individuals in the sample. For example, the retirement decisions of a sample of men and women would be expected to be more varied than those of a sample of men alone. Modeling the more varied decisions would require the inclusion of additional factors in one model or the development of independent models to explain the differences between men and women.

Table 3.3 summarizes how models treated sex differences among respondents. As table 3.3 illustrates, most of the models (27) were tested only on male samples. Only five models explicitly model women's retirement patterns. This omission can be traced in part to the data sources most frequently used in the model estimations. The original RHS surveyed only women who were single in 1969. Some information on married women is available for those single women who married by subsequent RHS data collections and for spouses of the married male RHS respondents. However, models based on these data lack the ability to generalize findings to a larger group of married women because the sampled women were not selected to be representative of any larger group. The NLS did not contain information on women of retirement age and the PSID surveyed heads of household, who were most often male. A second reason for the lack of models of women's retirement behavior is that women's work patterns are more varied than those of men and the factors which influence their retirement decisions are less well known, making modeling a more difficult enterprise for this group. Third, the very rapidly changing trends in women's work patterns mean that older generations' retirement behavior is not a reliable guide for the retirement behavior of the younger generation.

Table 3.3: Model Treatment of Sex Differences

Treatment	Total	Model Nos.
Modeled male behavior only	27	(Nos. 1,2,3,4,5,7,8,11,14,15,16, 17,19,20 21,22,24,25,26,27,28, 29,30,31,32,33,34)
Modeled female behavior only	2	(Nos. 12,18)
Modeled male and female behavior independently	3	(Nos. 9,13,23)
Modeled male and female behavior jointly	1	(No. 10)
Included sex as a predictor	2	(Nos. 6,35)

In addition to partitioning samples on the basis of the respondent's sex, samples were sometimes partitioned on other characteristics, such as the respondents' race and marital status. Table 3.4 summarizes how models treated race differences among respondents. Of the 20 models which reported how race differences were treated, 11 were partitioned by race but only two explicitly modeled the retirement decisions of black or non-white respondents. This does not mean that differences in behavior associated with the respondent's race were ignored in the remaining models. Rather many of them included both blacks and whites in the estimation sample and entered race as a predictor of behavior in the model specification. With regard to marital status, of the fourteen models which partitioned samples by marital status, the majority, 57 percent, used only married respondents in their tests. Two were tested on single and married samples; four, on only single respondents. Of these six, three were based on women who were single in 1969 but may have subsequently married. Of the remaining models, 3 percent included marital status as a predictor in the model. A summary of model treatment of marital status is provided in table 3.5.

Table 3.4: Model Treatment of Race Differences

Treatment	Total	Model Nos.
Modeled behavior of whites only	9	(Nos. 2,3,4,7,11,14,15,21,23)
Modeled behavior of non-whites only	0	
Modeled behavior of whites and non-whites independently	2	(Nos. 16,22)
Included race as a predictor	9	(Nos. 5,6,14,15,20,27,31,32,35)
No reported treatment of race differences	15	(Nos.1,8,9,10,12,13,17,24,25,26,28 29 30,33,34)

Table 3.5: Model Treatment of Marital Status Differences

Treatment	Total	Model Nos.
Modeled behavior of married individuals only	8	(Nos. 2,3,5,10,12,15,17,21)
Modeled behavior of unmarried individuals only	4	(Nos. 13[a],18,31,32)
Modeled behavior of married and unmarried individuals independently	2	(Nos. 16,23)
Included marital status as a predictor	16	(Nos. 1,4,6,7,8,11,13[a],14,19,20,26 27,28,29,30,35)
No reported treatment of marital status	6	(Nos. 9,22,24,25,33,34)

[a] Model no. 13 modeled the behavior of unmarried females and included marital status as a predictor of male behavior.

Many other characteristics were used as selection criteria for individual models to reduce the variability among respondents in the estimation samples. For example, several models excluded self-employed workers and/or federal workers. Others have excluded welfare recipients, farmers or men who have working spouses. One model disaggregated workers by the physical demands of their jobs and several, by health limitations. Reducing the sample variability in this way may simplify the model specification at the cost of increased generalizability. This loss in generalizability could be restored by testing the model on more than one of the partitioned groups and pooling observations when comparable results are obtained. However, this has not been done very frequently. In cases where the set of factors affecting retirement decisions and the nature of their effect are expected to differ across sub-groups, as is the case for men and women, blacks and whites, and more and less physically demanding jobs, different models are indicated. In these cases there is no trade-off between simplified specification and increased generalizability.

Predictors

One of the most important issues in modeling human behavior is the selection and measurement of predictors. The predictors are a set of variables used to describe different aspects of the sample population. Variation in the values for the predictors produces variation in the outcomes for different individuals or groups. Variation in outcomes not explained by variation in predictor values is considered to be error in the model.

One of two general principles typically guides the selection of predictors—theoretical or empirical validity. Some developers select predictors because they are consistent with some overall theory, most often economic life cycle theory, about retirement behavior. Others select predictors based on their observed relationships with, or their ability to explain, variation in the retirement decisions of workers. For some models, both principles are used to select the set of predictors—some are selected because of their theoretical validity and others because of their empirical validity.

The two principles are not always in conflict but they can be. For example, the replacement ratio of pension income relative to working wages may be more strongly related to the retirement decision (and thus have higher empirical validity) than the stream of future expected pension benefits. However, the latter variable is more consistent with life cycle theory than the former. A model developer concerned primarily with theoretical validity would choose the latter variable for a predictor despite its lower empirical validity. Forecasting experts believe that theoretical validity is preferable if a model is used to predict behavior as a consequence of policy change.

The specific predictors included in each model determine the types of policies that can be analyzed with the model. Most of the models include predictors related to the social security program. Many include other pension and income measures. A variety of additional predictors are included in the models.

Table 3.6 summarizes the social security related predictors in each model. The effects of social security on the retirement decision have been estimated in each model from one or more of the following predictors: current eligibility for reduced or full benefits, annual benefit amount and the change in benefit that would occur if retirement were delayed, the present discounted value of future social security benefits (social security wealth) and the changes in wealth that would occur with delayed retirement, the social security benefit to earned income ratio (the replacement ratio) and the change in that ratio that would occur with delayed retirement, the ratio of social security wealth to the present value of lifetime potential earnings, and predictors reflecting OASDI coverage (e.g., years of covered earnings, total covered earnings, and primary insurance amount).[4]

Table 3.6: Model Treatment of Social Security Effects

Treatment	Total	Model Nos.[a]
Eligibility for benefits	9	(Nos. 2,3,10,13,16,17,23,28,35)
Benefit amount	14	(Nos. 2,4,12,19,23,26,27,28,29,30 31,32,33,34)
Social security wealth (or a measure of future social security income)	17	(Nos. 5,7,8,10,13,15,16,17,20,21,22 24,25,26,27,33,34)
Replacement ratio	3	(Nos. 9,11,18)
Coverage	3	(Nos. 12,14,23)
OASDI covered earnings	1	(No. 23)
No treatment	2	(Nos. 1,6,)

[a]Models using more than one treatment are Nos. 4,10,12,13,16,17, 23,26,27,28,33 and 34.

The most frequent predictors are eligibility, benefit amount and social security wealth. Only six models do not include any of these three predictors. Two of these (Burkhauser, no. 1; Schmitt-McCune, no. 6) focused on non-federal pension acceptance and included no measures of social security effects. Three (Burtless-Hausman, no. 9; Gordon-Blinder, no. 11; and O'Rand-Henretta, no. 18) used replacement ratios to estimate the effects of social security and one (Gustman-Steinmeier, no. 14) used social security coverage.

A similar array of predictors has been used to estimate the effects of private pensions on the retirement decision. The most popular pension predictors are the wealth or asset value of the pension, current eligibility for benefits, coverage by or vesting in (having rights to benefit from) a private pension plan, and benefit amount or a proxy for benefit amount (e.g. years of service or contributions). Only one model used replacement ratios (the ratio of private pension benefits to earned income). Many of the models had no separate measure of private pension effects. This omission is largely due to information limitations in the data sources.[5]

---

[4] For complete structural models, where the direct predictors of retirement are parameters of labor-leisure preference functions, social security (and or private pension plan) rules are fully integrated in the determination of what alternative courses were available to the individual.

[5] See footnote 2.

The choice among benefit amount, wealth, and replacement ratios appears to be on theoretical grounds. Some proponents of life cycle theory argue that the wealth variable, which captures future benefit streams, is more appropriate under the assumption that it is the lifetime utility of working versus retiring that people maximize when they decide to retire. On the other hand, it could be argued that retirees are least likely to know their social security or pension wealth and therefore are less likely to use it for decision-making. Despite model developers' disagreements over use of the specific variables, studies of retirement decisions have found all 3 informative for explaining actual behavior. We were unable to locate any studies comparing the empirical validity of these three measures.

In addition to the availability and amount of retirement income, the models typically include other income or financial variables and non-income characteristics as predictors of the retirement decision. Examples of these variables are listed in table 3.7. The specific sets of predictors used in individual models are listed in the model descriptions in appendix III of the supplementary volume of this report. Generally the models appear fairly similar in the set of predictors.

Table 3.7: Predictors Used in Retirement Decision Models

Frequently Used[a]	Infrequently Used
Financial Predictors	
retirement income	subjective discount rate for future income
wage income	
non-wage income/assets	
income of spouse	
future income and assets	
Demographic Predictors	
age	geographic residence
race	sex
marital status	year of birth (cohort)
education	spouse's age
dependents	spouse's education
Work-related Predictors	
work experience	self-employment status
mandatory retirement provisions	employment sector
	occupation/industry
	job characteristics
	job attitudes
	local unemployment rates
	spouse's employment status
Other Predictors	
health	spouse's health
	subjective mortality
	available years of retirement
	retirement plans

[a] Predictor was used in more than 5 of the 35 models reviewed.

To study the sensitivity of model output to the selection and measurement of predictors, Gustafson (no. 16) developed a baseline model of the retirement decision in which he held all other parts of the model constant (the sample, the modeling process, the outcome variable and the predictor set) while he varied the measurement of single predictors. He focused on four critical predictors in the model—health, wages, social security and private pensions. His results demonstrated that outcomes from models of the retirement decision can be sensitive to differences in the measurement of predictors, especially social security and health. Thus, differences in results across models using a similar set of predictors may be due to differences in the way these predictors were measured.

Analytic Dimensions

Documentation

Documentation for this class of models consists of one or a few working papers, contract reports and/or professional journal articles. Reporting standards vary across disciplines (e.g., economics, sociology, psychology) and across document type but we found the documentation to be uniformly satisfactory for developing individual model descriptions for this review. More complete model documentation may exist in the model's computer code or elsewhere. We did not request or examine such sources. In some instances, developers noted that additional information on sensitivity analyses or results of alternative model specifications were available on request.

In the documentation we examined, there was not sufficient detail on sample selection, treatment of missing data, and measurement of predictors to allow independent replication of results, the most rigorous of reporting standards. For example, not all developers reported the discount rates and source of mortality rates used in the calculation of variables like social security and private pension wealth. However, most developers did provide elaborate detail on other aspects of predictor measurement. Although most developers reported model validity statistics, some did not. (Developer treatment of model validity is discussed in detail in the subsection below on validity.) In general, however, we were able to abstract from the documentation comparable descriptive information for all models.

Maintenance

The maintenance review dimension refers to the frequency and completeness with which models are updated and revised, or maintained for current use. On an individual model basis, there is little maintenance. Most of the models we reviewed were developed to serve a single purpose. The class of models (with a few exceptions) can be viewed in some respects, however, as a single model of life cycle theory which has been revised and extended by later-coming developers. For example, the most recent models take advantage of more current data, the most modern advances in calculating algorithms and computer technology, and include refinements in the measurement of predictors that were originally defined in earlier models. When the maintenance dimension is applied to the class of models, we find that the "model" has been frequently updated and revised by numerous experts and is continuing to be revised in this way. However, the decreasing availability of current longitudinal data sources for these models makes it difficult to update and maintain them.

Validity

Across all models, there is no widely accepted single measure of potential forecast error or standard measure for determining whether estimates of error are acceptable or not. For measures with known sampling distributions, there are standard conventions for evaluating the statistical significance of values. However, since virtually all models are based on very large numbers of observations, a prediction that accounts for a small amount of the difference among individuals could be statistically reliable at very high levels of confidence.

For the 35 models, even a cursory review of error was not possible due to lack of information. No developer reported an estimate of forecast error. For all but a few models, model validity statistics that indicate the models' ability to explain observed variation in outcome variables were reported but not emphasized. With one exception, there was no indication that developers had tested their estimated models on samples other than ones used to develop the models.

This dearth of information does not mean that developers were unconcerned with potential for error but it is expressed with caution (some developers do not recommend their models for forecasting purposes[6]) or a concern for theoretical validity. Most documentation included, for example, considerable discussion of the models' underlying theory of behavior, how the set of selected predictors and their measurement were consistent with that theory and how well model results for individual predictors conformed to theoretical expectations.

Although developers were concerned with theoretical sources of error in their models, most did not concern themselves with other sources of error, such as the reliability of the data and, in particular, the operational validity of the models. This is a general weakness for encouraging use of the models or their results for public policy analysis. In the remainder of this section, we show some desirable kinds of information on operational validity. These examples are taken from the few developers who provided readily interpretable statistics on their models' overall performance. They are presented in terms of the potential use of the model for explaining behavior, forecasting behavior in the absence of policy changes or forecasting behavior as a consequence of policy change.

---

[6] Models that developers explicitly stated should not be used for forecasting purposes include nos. 1, 8, 14, 20 and 23. We add current versions of models 19 and 33 to this list since the developers of these models did not find their results to be entirely satisfactory.

Assessing a Model's Ability to Explain Behavior

Of the 35 models we reviewed, the most variety of information on operational validity was provided for the initial specification of the Schmitt-McCune model (no. 6). The developers analyzed the role of various factors in explaining the retirement status of a sample of Michigan civil servants. Their documentation includes (1) statistics on the relationships between individual predictors and the outcome and among all of the predictors, (2) the internal consistencies (measures of reliability) of all predictors which were measured by more than one item, (3) the percentage of variance in the outcome variable that the model as a whole explained and that subsets of predictors in the model explained, (4) the percentage of the original sample that the model correctly classified on the outcome variable and similar percentages for subsets of predictors in the model, and (5) tests of the statistical significance for all but the internal consistency measures.

Some sample results from these analyses are informative. For example, a set of nine motivational psychological predictors explained ten percent of the variance in retirement decisions and a model based only on these measures correctly classified 66.4 percent of retirees and nonretirees. A set of nine demographic, work experience and income predictors predicted 22 percent of the same variance and correctly classified 73.5 percent retirees and nonretirees. Finally, the recommended model which included all of the latter predictors and 4 of the former predictors explained 28 percent of the variance and correctly classified 74.9 percent of retirees.

The percentage of variance explained by a model is a standard "goodness of fit" or model validity statistic for models using estimation procedures comparable to the one used by Schmitt-McCune. In addition to Schmitt-McCune, four models (nos. 3, 5, 24 and 32) used similar techniques as primary estimation methods and five (nos. 1, 8, 13, 16 and 23) used these techniques as secondary methods in conjunction with more preferred techniques. All four of the developers who used these methods as primary estimation techniques reported the percentage of outcome variance explained by their models. When these methods were used as secondary methods of estimation, two developers (nos. 1 and 8) reported model validity information only for their preferred technique, two (nos. 16 and 23) reported both preferred and secondary model validity statistics and one (no. 13) reported validity information for some but not all model versions.

Since these models are not at all equivalent, it would be misleading to compare results across models. However, it is useful to examine the range of validity outcomes. The Quinn model (no. 3) with 14 predictors explained 18 percent of the variance in labor force status of wage and salary workers; with eight predictors, 14 percent of the variance in labor force status of self-employed workers. The Mitchell-Fields model (no. 24) with two predictors explained 16 percent of the variance in age of private pension benefit acceptance for participants in ten plans. Within individual plans, their model explained from one percent to 31 percent of benefit acceptance age variance, with a median figure of approximately 10.5 percent. Finally, the Gohmann and Clark model (no. 32) with 13 predictors explained 31 percent of the variance in years to retirement after acceptance of social security benefits.

The percentage of correct classifications on the outcome variable is a standard statistic which is directly interpretable and is applicable to models in which the outcome represents membership in one or more categories. It can be informative, however, for other types of models. Only one developer (model no. 11) besides Schmitt-McCune reported the percentage of correct classifications. In lieu of correct classifications, one developer (no. 22) provided a comparison of the modeled and observed distributions of retirement from the labor force at various ages. Similar statistics would be useful for other models.

Other statistics provided by Schmitt-McCune, such as internal consistencies of predictors, intercorrelations among predictors and univariate tests of predictor-outcome relationships, are useful for independent evaluation of the appropriateness and role of individual predictors in the model. No other developer provided the first two sets of measures, although they are clearly appropriate for a few models.[7]

For models using techniques dissimilar to that of Schmitt-McCune there is less agreement on what standard validity statistics should be reported.[8] The most frequently reported model validity statistic in the models we reviewed was the model likelihood value or some function of that value. Twenty developers reported this model validity statistic. These values are not reported here because they are less directly interpretable than other measures. Unfortunately, most developers reported no model validity information beyond these values. Eight reported no model validity information at all. Information such as the percentage of correct classifications which was provided by Gordon and Blinder (no. 11) or a comparison of observed and modeled distributions of retirement across ages which was provided by Gustman and Steinmeier (no. 22) could potentially be provided for all of the models.

---

[7] Internal consistencies are appropriate when a predictor value is obtained by summing responses across two or more questionnaire items. Some of the health predictors used in the models were measured in this way. The internal consistencies of these predictors influence the confidence that is placed in results based on the predictors. When intercorrelations among predictors are used to generate final model solutions (and they often are) they also can be used to aid the interpretation of results and provide additional support for the validity of the model.

Many developers did report estimating their models with alternative techniques or on alternative samples drawn from a single data source. Some provided results from more than one estimation and others noted that such results would be available on request. Comparing results across estimations provides information on the sensitivity of the model to estimation technique and on the generalizability of model results to other samples.

It was more typical to find information on predictor validity than model validity. Virtually all of the developers reported the results of using alternative measures of predictors, provided validity information on constructed and imputed predictors, or presented and discussed other validity information on individual predictors. Much of the predictor validity information has been accumulated in numerous research reports that underlie the eventual development of the models we reviewed. Developers routinely cite this information as part of their validation of predictors.

In the context of forecasting, this focus is especially important for predictors such as social security, private pension and other future wealth variables since these predictors are forecasted from sample observations. All of these predictors require a forecast of future income, based on certain economic and demographic assumptions. This income is converted to present dollars, using present value procedures comparable to those described in chapter 2. Thus, our comments in that chapter on the importance of assumptions to the accuracy of forecasts apply to these predictors as well. Very few of the developers who used these predictors reported completely the sources and values of their economic and demographic assumptions.

---

[8] A good summary of the statistics that have been proposed and the problems associated with each for different types of models is provided in Maddala (1983).

Assessing a Model's Ability to Predict Behavior in the Absence of Policy Change

There are three ways in which a model's predictive validity or its ability to predict behavior can be assessed. One method is to use predictors measured on a sample of individuals at one point in time to estimate or predict outcomes that occur at a later point in time for the same individuals. A second method is to randomly split a sample into two groups, using one group to estimate the model, and then using the estimated model to predict the outcome in the second group. The third method is to use an estimated model to predict outcomes observed on an independent sample in a different time frame. The model can be used to either backcast prior outcomes or forecast future outcomes. This method typically requires more observational information than other methods. With all three methods, predictive ability is assessed by comparing predicted outcomes to actual observed outcomes.

The first method gives less independent evidence of predictive validity than other methods because the observed outcome is often used to estimate the model. This greatly enhances the odds that the model will be able to predict the outcome well. Nevertheless, of the three methods, it was the most frequently used in the models we reviewed.

Although the use of models to predict behavior is different from their use for explaining behavior, when predictive validity is assessed using the first method, the procedures or appropriate test statistics are identical to those used to validate the models' ability to explain behavior.

No developer reported validating their model with either the second or third methods.[9] Three models (nos. 15, 16 and 34) were used to produce backcasts (the third method) but not for model validation purposes. All three produced backcasts of the effects of past social security benefit increases. It is interesting to note that three different types of model—structural, longitudinal and non-longitudinal reduced form—concluded that social security played a minor, intermediate, and major role, respectively, in the early 1970s decline in labor force participation. This information could be used as part of the model validation (by using the divergence between backcasting predictions and actual outcomes as a measure of how well the model predicted these outcomes) although the developers did not use it in that way.

---

[9] Although documentation for the DYNASIM version of the Burkhauser-Quinn model (no. 13) did not include cross-validation information, it is likely that such information has been calculated by the model developers. This model was developed on a sample of respondents to the RHS. In the DYNASIM model it is applied to a sample of respondents to the CPS.

Assessing a Model's Ability to Predict Behavior as a Consequence of Policy Change

Appropriate procedures and test statistics for assessing a model's ability to predict behavior as a consequence of policy change are less well defined than for other model uses, although obviously when policy or other relevant changes do occur, actual behavior can be compared to what the models predicted would happen. In the absence of this kind of data, experts believe that correct specification of a causal model is a more trustworthy criterion for placing confidence in experimental predictions than is a model's ability to explain current behavior. Some current explanatory power is, of course, expected for all models. However, explanatory power alone is not sufficient to capture the effects of future policy change. Thus, both kinds of information are needed in order to evaluate the models.

Reviewing models on their theoretical validity was beyond the scope of this report. The issue surfaced in our classification of models by estimation methods. In that section we noted that the structural models of life cycle theory estimate individual labor-leisure preferences more directly than reduced form models of that theory. Theoretically, these preferences and their effects on decision-making are less sensitive to policy change than some factors (such as eligibility for social security) that are estimated by reduced form models. Thus, if all other judgmental criteria were equal across models, the structural models would be preferred over reduced form models for predicting behavior as a consequence of policy change. Good theoretical reviews of some of the models are available in Fields and Mitchell (1983) and Danziger, Haveman, and Plotnick (1981). In addition, a review of sources of theoretical specification error in some of the models is available in Gustman and Steinmeier (1983).

Summary

In this chapter, we reviewed 35 models of retirement decision behavior, largely models of decisions regarding labor force participation and drawing pension benefits. Most of these models were developed to estimate the relationship between social security and the retirement decisions of workers. Over one-third have been applied in the experimental analysis of public policy change. Some of these experiments have concerned retirement policy in areas other than social security. The major factors affecting model outcomes are specification of a theoretical model and selection or development of an estimation method for it, and selection of a data source and set of predictors.

The majority of models (32 of 35) were developed from the perspective of economic life cycle theory. Approximately two-thirds of these models (23) used reduced-form estimation techniques. Six modelers developed structural models of life cycle theory and four developed techniques to better estimate the longitudinal nature of retirement decision-making.

All but one of the models were developed on data collected prior to 1980. The majority were based on federally sponsored longitudinal data surveys, with 66 percent drawing on some data from the Retirement History Survey which was discontinued in 1979. Most of the models (27) were tested only on male samples. Independent models of the behavior of females and non-whites were rare: five for the former group and only two for the latter. In addition, many other characteristics were used in individual models to reduce sample variability and simplify model specification. This procedure introduced a loss in generalizability for the models.

The set of predictors varied widely across models. All but three included some social security related predictors of retirement, ranging from simple observations of eligibility for benefits to complex estimations of social security wealth that depend in large part on economic and demographic assumptions specified by the developer. In addition to social security, most models included a varied array of other income, demographic work and health-related predictors. A few models included unique predictors, such as attitudes, characteristics of spouses, and subjective mortality. Even for models using similar predictors, results can vary because of differences in how the predictor values are measured or estimated.

Documentation for these models was fairly uniform in content and level of detail. There was typically elaborate detail on the theoretical model, on methods of measuring or estimating unique or complex predictor values, and on the validity of individual predictors. There was less systematic treatment of sample selection, data quality, economic and demographic assumptions, and overall model operational validity.

Little on-going maintenance of individual models was found during this review. When model revision occurs, it usually results in a new model because either the theory or the methods of estimation are revised. When the maintenance dimension is applied to the class of life cycle models, the "model" has been frequently updated and revised. There is less promise on the availability of updated longitudinal data for future model maintenance. Major data series have been discontinued, thus jeopardizing the accuracy of forecasts depending on them.

Information on the operational validity of these models is seriously lacking. No developer reported an estimate of forecast error. There was no indication in model documentation that developers had tested their estimated models on samples other than ones used to develop the models and in general there was no discussion of the reliability of the sample data. Although virtually all of the developers reported some overall model validity statistic to reflect the model's ability to explain behavior, few provided information beyond this number. Three models backcasted behavior but did not use the results for validation purposes. Finally, we found no reports on the historical accuracy of any of the models.

Chapter 4

Models of Retirement Income

Models of retirement income are used to predict the future levels and distribution of that income. This chapter describes models which specifically forecast long-range income for elderly and/or retired persons.

We identified four major models, one of which has multiple versions, developed to forecast retirement income. These four models—DYNASIM, PRISM, MDM, and the AARP Age-Income Model—are computerized forecasting models that have been applied for public policy analysis, maintained since their original development and are currently available for use.[1] These models were developed by private contractors. Detailed summaries of individual models are provided in the supplementary volume of this report.

This class of models describes many aspects of the retirement income system, including characteristics of individual retirement behavior, of the labor market and of the programs which distribute retirement income. The primary focus of these models is on predicting income. In some instances, however, estimates of benefits paid out by a particular program are used to produce cost estimates. Some also make non-income predictions (estimates of population size, and labor market behavior, for example) which serve as input to other models.

Models of retirement income can be divided into two classes: (1) DYNASIM and PRISM, which use the individual as the basic unit of analysis (the microsimulation approach), and (2) MDM and the AARP Age-Income model, which use a group as the basic unit of analysis (the macrosimulation approach). The microsimulation models primarily estimate the distribution of income while the macrosimulation models primarily estimate future levels of income. Because of their size and complexity, and the amount of estimation they require, these models are the most speculative in nature of those we reviewed.

Background and Use

DYNASIM (Dynamic Simulation of Income Model) was first used at the Urban Institute in 1976. It was similar to an existing model, TRIM (Transfer Income Model) used for welfare policy analysis, in that it calculated the components of income for a sample of the population. It differed from TRIM in simulation technique, using dynamic rather than static aging. This technique made DYNASIM appropriate for making forecasts over a longer time period than was possible with TRIM. This longer forecast period also made DYNASIM more appropriate for retirement policy analysis. (Since TRIM has not been used for long-range forecasting of retirement income, it was not included in our review.[2]) Since 1976, several versions of the DYNASIM model evolved.

---

[1] Our assessment of availability was made in 1984 at the time of our data collection. According to HHS officials, MDM is not currently (1986) available for use. For details, refer to their letter to us which is reproduced in appendix II of this volume.

PRISM (Pension and Retirement Income Simulation Model), developed in 1980 under sponsorship of the Department of Labor and the President's Commission on Pension Policy, is similar to DYNASIM in its use of dynamic aging simulation; it has been applied exclusively to retirement policy analysis. MDM (Macroeconomic-Demographic Model) was developed in 1981 for the President's Commission on Pension Policy, and the AARP Age-Income Model of the Elderly was developed by Data Resources, Inc. (DRI) for the American Association of Retired Persons.

All of the models have produced baseline forecasts of future income for the elderly, although these forecasts are not readily comparable as they are made for different time periods, with different assumptions, and use different outcome variables. In addition, DYNASIM, PRISM, and the AARP Age-Income Model were all used to analyze the potential effects on income of various proposals for changing the social security program, including those adopted in the 1983 Social Security Act Amendments. The AARP Model has been applied to predict the effects of various proposed changes to Social Security cost of living adjustments. PRISM and DYNASIM have been used to estimate the effects of various proposed changes to the private pension system including mandatory universal pension coverage and the indexation of benefits to economic conditions.

In the remainder of this chapter, we analyze these four income models along four dimensions—outcomes, methods, data sources and predictors—and summarize model status on three analytic dimensions—documentation, maintenance and validity. We conclude with a summary.

Descriptive Dimensions

It is difficult to describe completely the large and complex microsimulation models. For example, the AARP Age-Income Model incorporates the DRI Macroeconomic Forecasting Model, a model with over a thousand equations, and hence contains numerous data sources, assumptions, predictors and procedures. Although the other three models are not composed of as many equations, they still involve a number of submodels and many items across these descriptive dimensions.

---

[2] For an evaluation of the TRIM model, see GAO (1977b).

The discussion in this chapter focuses instead on the items which can be readily summarized for this category of models: outcomes, simulation methods, and primary data sources. The predictors are obviously important components of these models and are discussed to a limited extent. More detail on predictors for each model is given in the supplementary volume of this report.

Outcomes

The primary outcome prediction for each of the 4 models is future retirement income. Each of the models makes predictions for a number of years into the future. The components of income which are predicted by each model are listed in table 4.1. Three of the models—PRISM, DYNASIM, and MDM—make calculations for seven sources of retirement income. The AARP Age-Income Model calculates overall income which includes some of these sources, but does not predict them separately. The three models that disaggregate income into its various components also can provide summary data for the entire population, such as the total social security benefits received in a particular year, a valuable feature for analysis of some social security financing issues.

Table 4.1 illustrates the variety of income information and the amount of detail each model can forecast for each year. All but the AARP model forecast social security, pension benefits, supplemental security income, and wages. Two of the models track Individual Retirement Account accumulations and distributions. MDM is the only model that forecasts Medicare benefits. PRISM and DYNASIM are the only models which calculate taxes for the purpose of determining disposable income.

Table 4.1: Available Income Breakdown for Models of Retirement Income[a]

Income Source	Models		
	DYNASIM	**PRISM**	**MDM**
Social Security	retirement, disability, survivors, and dependents benefits are calculated	retirement, disability, survivors, and dependents benefits are calculated	retirement, disability, survivors, and dependents benefits are calculated
Private Pension	benefits are calculated based on an assignment of pension plan characteristics representative of private plans	benefits are calculated based on assignment of a pension plan from a sample of plans	an average benefit is calculated for individual age-sex groups for defined benefit and defined contribution plans
Public Pension	benefits are calculated based on an assignment of pension plan characteristics representative of public plans	all federal employees are assigned to the CSRS; state/local employees are assigned to social security integrated plan using a method similar to private pension plan assignment	an average benefit is calculated for individual age-sex groups for seven categories of public employees[b]
Supplemental Security Income (SSI)	calculated	calculated	calculated
Individual Retirement Accounts (IRAs)	an IRA (or Keogh) is calculated based on projected coverage rates; distributed evenly across retirement years	an IRA (or Keogh) is calculated based on projected coverage rates; distributed evenly across retirement years	planned revisions to the model include adding IRAs
Wages for Working Elderly	calculated	calculated	calculated
Taxes	federal FICA and income taxes are calculated	federal and state income and FICA taxes are calculated	not calculated[c]
Medicare Benefits	not calculated	not calculated	calculated

[a] The AARP model does not disaggregate income by source.

[b] The categories are: Federal Civil Service, Military Enlisted Persons, Military Officers, State Local Hazardous, State Local General, State Educators, Local Educators.

[c] FICA taxes are calculated by the model but cannot be used to make adjustments to income.

PRISM, DYNASIM, and the AARP Age-Income Model can all produce forecasts of total income. For the two former models this is done by calculating the sum of each of the predicted income components for each individual. The AARP model only forecasts total income. It is not clear from the documentation what all the components of the AARP total income are, but they do include wage income and social security income.

It is possible to sum all of the MDM aggregated income components. For example, the total income for males aged 62-64 could be calculated from each of the predicted components (all of the social security benefits, pension benefits, SSI benefits, etc.). Because of the aggregated nature of the model, however, it would not be possible to determine which people are receiving which components of income in order to determine how well off the population is.

Aside from the income variables, three models (DYNASIM, PRISM and MDM) produce estimates of other factors, including various predictions on the future demographics of the population and some forecasts of more general economic trends such as inflation rates, unemployment, etc.

With regard to subgroup analysis, table 4.2 lists different population sub-groups which can be described for each of the models. The basic unit of analysis affects the ability of the models to disaggregate according to various demographic cohorts. By tracking individual level information, two models (DYNASIM and PRISM) can report income according to numerous population characteristics, the limit being the number of individual characteristics available for the initial population and predicted by the model. The other two models (MDM and AARP) do not track individual level information and therefore cannot disaggregate for many characteristics other than age and sex. Disaggregation by these two characteristics is possible because the two models make separate forecasts for specific age-sex groups.

Table 4.2: Available Demographic Breakdown for Models of Retirement Income

	Models			
	DYNASIM	**PRISM**	**MDM**	**AARP**
Basic Unit Of Analysis	family units, (information is recorded for individuals as well)	family units, (information is recorded for individuals as well)	groups of individuals in selected age ranges	groups of individuals in selected age ranges
Demographic Categories				
Age	any individual age or any interval	any individual age or any interval	55-58, 59-61, 62-64, 65-67, 68-71, 72+	55-61, 62-64, 65-71, 72+
Sex	disaggregated	disaggregated	disaggregated	disaggregated
Household status	marital staus, age of children, number of children	marital status, age of children, number of children	not disaggregated	single versus a consumer unit of 2 or more members
Occupation	ten industry classifications[a]	eleven industry classifications[b]	employment sector for recipients of different types of public pensions (see table 4.1)	not disaggregated
Education	number of years of education and highest level of education (grade school, junior high school, college and graduate school)	some educational information for the original population to age 25[c]	not disaggregated	not disaggregated
Race	white or nonwhite	race information for the original population[c]	not disaggregated	not disaggregated
Other	both of the microsimulation models have other information (e.g., disability status, years on current job)		none	none

[a] The ten industry classifications are agriculture, construction and mining, manufacturing, transportation, utilities and communication, trade, finance, insurance, real estate services, state and local governments and federal government.

[b] The eleven industry classifications are agriculture, construction and mining, manufacturing, transportation, trade, finance, insurance, real estate services, self-employed, state and local government, federal government.

[c] PRISM uses a baseline population of 25-64 year olds. Education is known for the initial population and the model assumes no additional education beyond age 25.

[d] Although the model does not use race as a predictor for explaining behavior, it uses the same initial CPS as DYNASIM which contains information on an individual's race.

Methods

Macrosimulation models (MDM and AARP) which forecast income for the elderly (retired and non-retired) population require a sub-model of the United States economy to estimate future economic factors. Each also contains a sub-model for projecting the size and composition of the future population. Demographic and economic output which is disaggregated for population sub-groups is used to calculate income.

Although they use the same general approach, the two models identified in this category differ in their implementation of that approach. One major difference in approach is how the macroeconomy is described. The AARP model is based on the DRI quarterly forecasting model of the macroeconomy. The MDM model is based on the Hudson-Jorgenson model of the macroeconomy, a model designed to forecast long-term growth as opposed to short-term business cycles. Other differences arise because the AARP Age-Income Model is, in general, a more aggregated model than MDM.

Microsimulation models (DYNASIM and PRISM) take advantage of the diverse characteristics and behavior of individuals in order to describe differences across groups of those individuals. First the future behavior of each individual in the sample population is predicted and then the results for groups of those individuals with a common set of characteristics are aggregated. This differs from the macrosimulation approach which directly predicts the average group response.

In general, microsimulation models can be characterized by their aging technique. "Aging" refers to the way in which the model projects the base year population to some future year. Both of the microsimulation models discussed here, DYNASIM and PRISM, use "dynamic aging." Dynamic aging models simulate the changes in the population (birth, death, migration, etc.) year by year from the base year through the future year. The alternative technique, "static aging," does not attempt to construct the population each year, but instead uses external predictions to reweight the initial population to reflect those predictions. Static aging models are used primarily for short range projections.

Dynamic aging models simulate events (e.g., marriage, job change, retirement) and conditions (e.g., industry of employment, wage, pension coverage) for every individual in a sample of the U.S. population over a specified period of time. This is done through application of a probabilistic technique called Monte Carlo simulation. The application begins with a record for a given year for a given individual in the population which describes various characteristics of that individual. Next, the data for that individual are exposed to the first module in the model which might be, for example, the mortality module determining whether an individual will die in a given year.

If probabilities indicate that the individual would not die in that year, then the data for that individual would be exposed to other modules (e.g., childbearing, job change, retirement, etc.) in the model. For each of the remaining years of the simulation, the data for that individual would go through a similar process. The results in each year will not necessarily be the same because the individual's characteristics (which determine the probabilities) are changed by the model, and because outcomes are random.

All individuals in the population are processed in this way so that the result of the simulation is a longitudinal record for each individual in the population and cross sectional group results for individual years, with the last year usually of most interest.

To account for the large potential for error in predicting these many individual behaviors, dynamic microsimulation models adjust their results by constraining them to external aggregate predictions (on employment, for example) which in many instances are produced by macrosimulation models. The focus is on describing diversity in individual behavior, rather than final aggregated income levels.

PRISM and DYNASIM both rely on the microsimulation technique although they differ in the methods used in each of their modules to predict behaviors.

There is some effort now to join a microsimulation and a macrosimulation model so that they can be simulated together. It has not been successfully implemented yet, although there are efforts under way at the University of Michigan to link an annual version of the Michigan Quarterly Econometric Model of the U.S. economy (MALTHUS) to a version of DYNASIM (MASS). This would allow feedback between micro and macro responses. Individual results from a microsimulation model in a given period would be aggregated; this information would be fed into a macrosimulation model to generate macroeconomic results (total output, investment, etc.) which would be used to constrain the microsimulation output in the next period. The simulation would continue period by period. This technique would account for individual behavior, macroeconomic forces and the interaction between them.

Data Sources

The AARP Age-Income Model uses annual Current Population Survey (CPS) data to estimate the income distributions for the different demographic cohorts in the model. Although many other data sources are used, the CPS data are important as the basis for estimating one of the key assumptions in the model.

The two microsimulation models, DYNASIM and PRISM, simulate the life experiences of an entire population, and hence the initial population is a key feature of these models. Both models extract their populations from CPS data which have been matched to social security earnings records for individuals in the survey. DYNASIM uses the March 1973 CPS-SER Exact Match File and PRISM, the March 1978 CPS, matched to social security earnings histories and updated with information from the March 1979 CPS and the May 1979 CPS Pension Supplement. Other data sets could be used. The DYNASIM model was simulated at one time using a population from the Panel Study of Income Dynamics. None of the data sets contain all of the necessary information to simulate these two models. Thus both models rely on various imputation procedures to assign "missing" characteristics to individuals in the population. MDM bases its initial population on 1980 Census figures.

All of the models use numerous additional sources of data including forecasts from macroeconomic models of the national economy, forecasts developed for the OASDI cost estimate models, forecasts from other sources, and multiple data sets. Many of these sources are identified by model in the supplementary volume of this report.

In earlier chapters, we discussed the importance of reliable and accurate data to reducing the potential for forecast error. For the models reviewed in the present chapter, this is an even more critical issue because of the sheer numbers of such sources.

Predictors

Identifying and summarizing all of the predictors for the income models was not feasible in our review. These models contain numbers of submodels and, within submodels, numerous equations are estimated with a variety of techniques and differing sets of predictors. Many of these equations are estimated with techniques similar to those described in chapter 3 for retirement decision behavior models. Our discussion in that chapter of the role of predictors in determining model outcomes applies to each of the behavioral equations estimated similarly in the income models. Separate treatment of predictors would be warranted for equations estimated with other techniques. Summaries of key predictors of outcomes, which are largely demographic characteristics (e.g., sex, age) and work history variables, are provided by submodel within model in the supplementary volume of this report.

Analytic Dimensions

Documentation

Documentation for each of the models of retirement income is substantial, relative to models in the other categories. Documentation also is fairly current; for three of the models, documentation was current in 1984: PRISM (February 1984), MDM (June 1984) and AARP (September 1984). The documentation for the Urban Institute's version of DYNASIM (December 1982 and November 1983) is slightly older. Individual users of the various other versions of the model do not publish documentation on their changes to the model.

The AARP model documentation presents the most detail with statistical output from estimated equations: parameter estimates, some validity measures, and a graphic presentation of the actual and predicted values for each equation. The documentation does not contain, however, a description of how model simulations are performed; documentation for the other models does contain such descriptions.

One of the models, MDM, has a user's manual available. It was developed by the National Institute of Aging and is intended as a guide for use of the model on the National Institutes of Health computer system.

In general, the documentation for these models provides useful information on how the models operate, but much information that might be of use to a model evaluator still is missing. For example, the specifics of model simulation and detail about use of the various data sources is missing in part for all of the models. In short, the documentation does not provide enough detail for replicating model results or independently testing validity.

The size, complexity, and evolution of these models may make it difficult for developers to maintain complete and current documentation. The documentation that is available is useful for understanding the models and sub-models, but is not sufficient for potential model evaluations.

Maintenance

Maintenance and update activities appear to be related to model use. The AARP Age-Income Model is the only model of the four reviewed in this chapter which is maintained and updated on a regular basis. The

AARP model is the only one that produces regular forecasts, and maintenance activities precede the annual forecast. The other 3 models do not produce regular forecasts and are updated in connection with specific model use, although minor revisions to these models appear to be going on continuously.

Validity

There is little published information on the operational validity of these models. DYNASIM and the AARP model documentation report some validity measures for the estimation of some model equations. MDM documentation includes the results of backcasting several outcomes for the period 1970-1979. No information is available, however, on the potential for forecast error in final outcomes for any of the models. Developers reported to us that they monitor the accuracy of their assumptions, calculate validity statistics on estimated equations and perform sensitivity analyses. However, the results of all these analyses are not routinely published. (The AARP model documentation includes validity statistics on some estimated equations.)

Although it is not possible to test the long-range forecast accuracy of these models which have been making forecasts only recently, critics of DYNASIM and PRISM note that other validating steps such as backcasting and sensitivity analyses have not been tried, or if tried, not reported.

The one formal comparison of DYNASIM and PRISM was conducted by Haveman and Lacker (1984). The two forecasts which they compared are reproduced in table 4.3.[3] For these baseline forecasts of both models to the 21st century, they found considerable discrepancies. The suggested reasons for these discrepancies include: differences in the initial population samples, different specifications of the relationships represented in the models, use of different data sets to represent those relationships, different judgments in the absence of data, and different assumptions. They were unable to establish which factors were responsible for the forecast discrepancies because the extensive sensitivity testing necessary for such a conclusion was beyond the resources of their project. Thus, there is little basis for deciding which forecast to use.

---

[3] These forecasts were not made with the same set of assumptions. Some differences in recipiency rates may be due to the fact that DYNASIM results are for married and unmarried individuals, while PRISM results are only for unmarried individuals.

Table 4.3: Comparison of Projections From DYNASIM and PRISM

	DYNASIM[a]	PRISM[b]	DYNASIM	PRISM	DYNASIM	PRISM
Males	1982	1985	2000	1995–05	2020	2015
OASI benefits[c]	$5,084	$4,401	$5,573	$5,733	$7,865	$7,875
Private pension benefits	$1,876	$3,903	$3,509	$6,160	4,521	$7,438
Percent receiving private pension	31.1%	29.3%	54.1%	48.5%	60.5%	49.3%
Females						
OASI benefits[c]	$3,115	$3,002	$3,452	$3,992	$4,808	$5,532
Private pension benefits	$846	$2,321	$1,584	$2,287	$1,897	$3,756
Percent receiving private pension	11.2%	11.7%	24.3%	30.2%	40.6%	46.5%

[a] DYNASIM projections are for 65-67-year-olds.

[b] PRISM projections are for 65-year-olds.

[c] All dollar figures are average annual benefits in constant 1978 dollars.

Source: Haveman and Lacker (1984), p. 4.

As a proxy for the costly in-depth evaluation, they qualitatively assessed the models, sector by sector, pointing out differences in approach, the theoretical validity of the approach, and the potential impact of the differing approaches on outcomes. For example, the two models differ in how they assign pension plans to eligible individuals in the sample population. PRISM assigns an actual pension plan to those individuals from a sample of actual pension plans. DYNASIM, on the other hand, constructs a pension plan by assigning a set of pension plan characteristics from a universe of those characteristics. The authors observe that DYNASIM's limited set of characteristics may not accurately capture the diversity of actual plans and suggest that PRISM's approach may be better, pointing out that DYNASIM captures the influence of many demographic factors in predicting labor market behavior, and PRISM, while deemphasizing the number of influencing factors, concentrates on accurately depicting the intertemporal pattern of an individual's labor market behavior.[4] They conclude, however, that it is not possible to determine which model is structurally superior, due to the lack of validity testing, especially sensitivity analysis and backcasting.

MDM documentation includes a comparison of some forecasted outcomes with similar forecasts made by the OASDI model (discussed in chapter 2) and the Bureau of Census, with discussion of reasons for some of the differences between forecasts, such as use of different assumptions.

---

[4] Others have been critical of the PRISM approach, suggesting that the survey of plans from which the model chooses plans may not represent the universe of plans appropriately.

Some of the similarity between MDM and OASDI forecasts can be attributed to MDM's use of administrative data which is also used in the OASDI model, its direct use of OASDI model forecasts (e.g. the future number of OASDI secondary beneficiaries) and the use of comparable methods for some outcomes (e.g. future payroll tax payments). No comparison of MDM forecasts with those of other retirement income models is available.

The validity of income models is a key issue. Their complexity and long range forecasts make them highly susceptible to error. Uncertainties about the magnitude of error makes it difficult to interpret the forecasts of these models.

Summary

In this chapter we reviewed four models of retirement income: DYNASIM, PRISM, MDM and the AARP Age-Income Model. All four models have been used for retirement policy analysis. DYNASIM and PRISM use microsimulation methods with dynamic aging to forecast the distribution of retirement income across various segments of the population. Outcomes from both models can be disaggregated by similar demographic characteristics, and total income forecasts can be broken down for both models into income sources. MDM and AARP use macrosimulation methods to forecast future retirement income levels. Of these two models, only MDM forecasts are broken down by retirement income components and only AARP forecasts are of total income. Both models can produce demographically disaggregated outcomes.

Each model represents the complex interactions of a number of submodels and equations and as such requires numerous assumptions, input data, and predictors. Many of the assumptions are derived from other models, such as macroeconomic models of the national economy, and used as input data to the income models. While each model uses a variety of external data sources, data collected by the Census Bureau is of central importance to all four. DYNASIM and PRISM extract their initial population life experience data from 1973 and 1978, respectively, Social Security-matched CPS files. MDM bases its initial aggregate population on 1980 Census figures and AARP uses annual CPS data to estimate income distributions.

Predictors are numerous and vary depending on the particular component of behavior or income being predicted. However, key predictors for all models are largely demographic and work history variables.

Documentation for each of the four models was substantial and fairly current, although federal versions of DYNASIM have not been independently documented. The specifics of model simulation and detail about use of the various data sources was missing in part for all of the models.

Major model maintenance appears to be related to specific model applications with the AARP model undergoing more frequent updates in connection with its regular (annual) use. The extent of revision varied across models.

As with other model categories, there is little published information on the operational validity of these models. No information is available on the potential for forecast error in final outcomes for any of the models. Developers reported to us that they monitor the accuracy of assumptions, calculate validity statistics on estimated equations to assess their explanatory power, and perform sensitivity analyses. However, the results of these analyses are not routinely published. Although the models are too new to test their long-range forecast accuracy, other validating steps, such as backcasting or cross validation, are possible, as MDM documentation shows, but either have not been done or are not reported for the other models.

Chapter 5

Conclusions, Matters for Consideration, Agency Comments and GAO's Response

In the previous chapters we identified and reviewed 32 models of federal retirement program costs, 35 models of retirement decision behavior and 4 models of retirement income. Our reviews focused on descriptive dimensions of models that influence the forecasts that models produce. These dimensions were outcomes, methods, data sources and predictors. We also summarized what kinds of public documentation are currently available for the models and the kinds of information that documentation contains, what provisions model developers have for updating and maintaining models and what efforts they take to monitor potential sources of forecast error. In this chapter we summarize these reviews and discuss their implications for policymakers and model developers.

Summary

In the preceding chapters, the sources of forecast error for cost, behavior and income models were presented. The predominant source of forecast error for the cost estimate models is the economic and demographic assumptions they use. They are also sources of forecast error for retirement decision models. However, more important sources of error in decision models are the survey data on which they rely and the identification and estimation of predictors. All of these factors are also sources of error for the income models, but they operate in multiplicative fashion because of the number of events the models forecast. Because of the increasing opportunity for forecast error across model classes, the assessment of that error is also increasingly complex.

In conducting our review of these models we planned to describe the results of model evaluations that have been performed by others. Model evaluations intensively examine models on a variety of descriptive and analytic dimensions. We found a noticeable absence of such model evaluations for the models we examined. This absence is notable given the importance of model evaluations for determining the overall quality of the models and the credibility of modeling outcomes.

We also noted the virtual absence of publicly available information on operational validity for most models in all three categories. We found no estimates of the potential for forecast error in any model and no reports on the historical accuracy of forecasts for 70 of the 71 models. In the documents we examined, there was little systematic treatment of the issue. Results of sensitivity analyses were reported for a few of the cost and many of the retirement decision models but explanatory validity was treated for only a handful of retirement decision models and for some equations in a few of the income models. The results of backcasting outcomes were reported for only one of the income models and 3 of the decision models. The limited information we did find was so incomplete that even a rough assessment of model quality was not possible. Since all of the cost and income models and at least one-third of the behavior models have been used for public policy analysis, this means that policy-makers may be basing decisions in part on forecasts with unknown validity and unknown potential for error.

Related to the absence of information on model validity was an absence of other critical information in model documentation. For example, an analysis of actuarial gains and losses which could provide some useful information on model assumptions is not a standard feature of cost model documentation. Critical components of the OASDI cost estimate model were not documented at all and CSRS model documentation, although more complete than for most cost models, was less complete than that for the OASDI and Military Retirement models. Documentation for behavior models included little treatment of data quality and generalizability issues, and documentation for income models, though substantial, also omitted quality-related information. Overall we found documentation focused on either process or outcomes with little self-assessment on credibility. Several developers reported to us that they engage in self-assessment activities but the results are rarely published.

Model maintenance proved a relevant dimension of review for models of retirement program costs and retirement income. The models of retirement program costs we reviewed all produce annual forecasts on a regular time schedule. These models are also annually updated and revised to some extent to reflect changes in the law, changes in the covered population and changes in assumptions, particularly economic ones. Some models are also revised outside of the regular maintenance cycle. The models of retirement income we reviewed are updated and revised periodically, but not regularly. Revisions tended to be made for new and specific applications. Exceptions include the AARP Age-Income Model which produces annual forecasts and is updated and revised annually. All four of the models were revised to include the 1983 legislated changes in the social security program and more current baseline data. Models of the retirement decision were not reviewed along the maintenance dimension because they were not designed for periodic use. In some sense, the entire class of retirement decision models could be viewed as revised, extended, or alternative versions of a single life cycle model. If the class were viewed in this way, we would conclude that the model has been frequently updated and revised to take advantage of more recent data and the most recent theoretical and computational advances. The majority of these models, however, relied on data from a federally sponsored survey—the RHS—which was discontinued in 1979. The decreasing availability of current and relevant public data for use by these and the income models may present a maintenance problem for them in the future.

Descriptive information on the models' outcomes, methods, data sources, predictors and predictor values/assumptions was available in part for all of the models. Within each category of models, the level of detail on a particular topic was fairly uniform. Despite this consistency, we encountered difficulties in interpreting some of the descriptive information. The major problems for cost models were the absence of a standard used nomenclature for actuarial methods and a lack of clarity in the reporting of certain assumptions. As a consequence, ambiguity remains about exactly how the cost model forecasts are produced. For both retirement decision and income models, there was virtually no treatment of data quality and reliability. Thus, it was difficult to determine the extent to which modeling results were based on observed as opposed to constructed or imputed information and whether and to what population modeling results could generalize. Although there was missing detail in all model documentation, we found the above problems most critical.

Our review found that while models exist for all three outcomes, with considerable effort in development and maintenance, users of model forecasts are at risk from several sources. First, there is a serious lack of published information on the operational validity of the models. Their use rests on faith in the developers' attention to error reduction, but the user has little help in selecting the model or interpreting the results on the basis of readily available information about forecasting error. Second, documentation for some key models (such as OASDI) is insufficient to know what choices have been made in judgment-call variables which can notably influence the forecast. Third, for some models the lapses in or discontinuation of essential data sets means that projections are based on antiquated data—for example, retirement decisions of the labor force in 1969, which we already know was different in composition with regard to gender and may be different in other variables affecting retirement from the labor force of the mid-1980s.

Conclusions

Because of the speculative nature of forecasting, it would be helpful for model developers to demonstrate to potential forecast users that their work is credible. A first step toward achieving this goal is to invest more effort in model validation than is currently being given. Such effort might include developing methods to estimate forecast error, developing professional standards for model validation and increasing the amount of documentation currently allotted to model validation. We believe retirement decision and income models could be improved if information on data quality and data generalizability were included in the published documentation. Information about program cost models would be improved if full scale evaluations were performed and the results publicly documented. These efforts to validate and document models more completely could increase model developers' staff time and other costs.

Models sponsored and used by federal agencies could benefit from similar initiatives for increased validation and evaluation. Special consideration of the data needed for maintaining the retirement decision and income models may be useful when planning continuing and future data collections.

Matters for Consideration by the Congress

In view of the information deficiencies we have reported, the Congress may want to consider whether additional guidance to those federal agencies sponsoring or supporting retirement forecasting models would be helpful. Error-free forecasts are not possible but the Congress may want to consider whether the information on error potential that is now available is adequate, given the importance of forecasts in setting retirement policy. For example, because the federal government is a major sponsor/user of models of all three outcomes, it may be useful to establish federal documentation standards for retirement forecasting, and perhaps other models developed for the federal governments' use by employees or through contractual arrangements. Such standards might include a requirement that validation analyses be documented.

Agency Comments and GAO's Response

The Department of Defense, the Department of Health and Human Services, the Department of Labor and the Office of Personnel Management were invited to review and comment on GAO's analysis of retirement forecasting models. All responded and copies of their official comments are reproduced in the appendices. In general, these agencies which have primary responsibility for the matters discussed in our report agreed with our overall conclusions that efforts to improve model documentation and validity and to ensure that current data are available for model use would result in increased model/forecast quality. Beyond this general area of agreement, the agencies provided numerous specific comments. We corrected several places in the text where agencies pointed out errors. Some comments did not cite errors, but offered elaboration or discussion of points we raised. And some comments noted disagreement with statements in the report. The comments and our response to them are summarized below by agency.

Department of Defense (DOD)

DOD did not raise any question about the general analytic criteria we used to review program cost models or our specific description and summary of the Military Retirement System model. Their major criticism of the report is the source we used to define actuarial methods. We state in the report that there is no standard nomenclature for methods in use. This statement is a revision of the statement made in the draft reviewed by DOD that there is no standard. There are several standards, including the one we used to define actuarial methods. We interviewed practicing actuaries and found multiple standards in use, a finding that is consistent with our own experience in reviewing actuarial reports. In our view, the source we used to define methods is the easiest for non-actuaries to understand. Our use of the source does not imply a recommendation of it as a standard for future use. The problem DOD noted concerning the classification of the frozen initial liability method is an example of our point that the use of different nomenclatures across actuaries makes it difficult to interpret exactly what method was used in a given forecast. DOD incorrectly attributed to us a conceptual error concerning our discussion of accrued benefit methods. We agree with their point but did not say that accrued benefit methods cannot have an actuarial liability. We said in the draft, as we do in our report, that the accrued benefit methods are all with actuarial liability methods by definition.

DOD also expressed concern that readers may misunderstand our discussion of the assessment of forecast accuracy for retirement cost models. We agree with their point that long-term forecasting models should be evaluated on the basis of their long-term performance and refer the reader to our discussion of this issue in chapter 2. We also agree that the analyses DOD performs are good validity indicators. They are not, however, evaluations of forecast accuracy.

DOD noted that we did not mention their models of retirement decisions. This was a consequence of our stated objective to survey and review only models of civilian retirement decision-making.

Department of Health and Human Services (HHS)

HHS provided some general observations about our report and several specific comments. Their general comments reflect agreement with our view that sound retirement-related forecasts are important for national policymaking and that documentation and validation of retirement models are important aspects of model quality. Overall, they viewed the report as a valuable reference for the modeling field. They noted that since the completion of our data gathering in 1984, they have taken actions to update and improve documentation and validation of the models they work with. Their major criticisms of the report were that it (1) inappropriately evaluates all types of models by a standard set of criteria; (2) overlooks past contributions of models used to assess Federal operations; and (3) gives insufficient attention to the constraints in time and money faced by model developers, to the difficulties involved in attempting to improve a model's forecasting ability, and to the limited improvement they believe is possible in light of the effects of unanticipated changes in economic and demographic conditions.

With respect to HHS' first general comment, on the criteria we use in evaluating models, we disagree that they are either arbitrary or inappropriate for evaluating models of different sizes and purposes. The criteria are general, drawn from literature in the field, applicable to all types of models as we discussed in our 1979 publication Guidelines for Model Evaluation, and are tailored in our discussions in this report to the purposes of each type of model. We acknowledge in the report that there is disagreement over accuracy, among other dimensions of operational validity, but we believe accuracy is important to assess in an inventory aimed at policy-makers who may use models to aid decisions. HHS particularly criticizes our evaluation of retirement decision models and suggests a more useful analysis would compare and summarize the findings from this type of research. We believe the differences among these models make such a comparison very difficult and such a comparative synthesis was not our purpose in this report.

Concerning the second general HHS comment, that we omitted the beneficial past uses of models, we acknowledge that they have been widely used, but to assess their historical contributions was beyond the scope of this inventory of current models.

With respect to the third major HHS concern, we agree that there are time and financial constraints as well as other difficulties that model developers face which may reduce the amounts of attention they give to assessing and documenting model performance and other aspects of the modeling endeavor which can affect model performance. We did not systematically study time and financial constraints for the present report nor did we analyze the trade-offs involved in devoting more resources to model assessment and documentation. We expect that these constraints and trade-offs vary widely and believe their assessment should be done on a model by model basis. For this reason, our conclusions only suggest actions that model developers and sponsors might take. We make no recommendations in this regard. In the report we do discuss some of the difficulties involved in validating model performance. For example, in our discussion of the long-term forecast accuracy of cost models (see chapter 2), we note that the time lag between the forecast and the actual experience may be so long that there is little practical interest in how accurate the forecast was. In chapter 3, we discuss different problems associated with validation of the decision models. Despite the existence of these problems, there are appropriate intermediate analyses of factors which can affect forecast accuracy that could be performed, but in general are not.

Finally, we disagree with HHS that only limited improvement is likely. In fact, because of the missing information on models' current performance that we document in our report, the degree of potential improvement is impossible to know at this point. And since even small errors can have large long-term consequences in forecasts spanning many years, their correction can be important.

The remaining HHS comments are specific to the OASDI cost model reviewed in chapter 2 and the MDM income model reviewed in chapter 4. These comments and our response to them are summarized below.

First, HHS disagreed with our conclusion that documentation for the OASDI model was incomplete and that the accuracy of the short range projections of the OASDI is not evaluated. They cite two document sources which summarize the long-range model that were available during our review and note that descriptions of the short-range model are included in the OASDI Trustees Report, beginning in 1986. With respect to evaluation of forecast accuracy, they note that the Trustees Report has regularly included a comparison of the most recent actual experience with outcomes forecasted in the two previous years.

In response, we believe our report accurately describes the status of documentation and validation for the OASDI model. As part of our review, we examined the two publications mentioned by HHS. Neither publication includes a summary description of the method used in the long-range model to estimate revenues for the first 10 years of the forecast horizon or a description of the short-range model. We note in our report that plans were underway at the time of our review to document the short-range model and are pleased to see that a description of it is now available to the public. We also are aware of the comparisons between actual and forecasted outcomes provided in the annual Trustees Reports. Although this information is an indicator of model validity in the same way that analyses of actuarial gains and losses are, we do not view it as an evaluation of the forecast accuracy of the short-range model because even simple findings that might be expected from such an evaluation are not provided. For example, there is no way to tell from the information provided if short-range forecasts made in one year decrease in accuracy as the time between the forecast date and the future experience increases. This is because the information provided in the Trustees Report compares the most recent actual experience with forecasts made in two different years. There is also no way to tell if the most recent forecasts of the current year's experience are more accurate than ones done farther in the past because only two prior forecasts are provided for comparison purposes. Thus, although the cited information is a useful indicator of recent accuracy, we believe it is incomplete for evaluation purposes. Our conclusions in this regard are actually supported by other comments in the HHS review letter. They describe briefly how a proper evaluation of a model's accuracy should be done and then explain why such evaluation has not been feasible. For these reasons, we believe our report accurately summarizes the status of the OASDI model at the time of our review with respect to both documentation and validation.

Second, HHS disagreed with our assessment of information that is available on the operational validity of the MDM model of retirement income and our overall summary of available information for the class of income models. With respect to MDM, they mention various kinds of information that are available. We have amended the report to include discussion of the information available in the cited document. We did not change our discussion concerning unpublished documentation. That is, we state that developers reported that they validate model equations but do not routinely publish the results. MDM developers are included in that statement. We also made no change in our overall summary of what information is available for the class of models except to note that the appropriate analyses of backcasting results, done for the MDM model, could be done but is not for other models.

Department of Labor (DOL)

DOL believed our report would be a considerable service to researchers and practitioners and agreed with our emphasis on improved data for use in models. The Department in most of its comments did not suggest that our report needed correction, but characterized its specific points as observations with the purpose of furthering discussion of pension cost estimating methods.

For example, concerning retirement cost models, DOL stressed the difficulty of measuring a component of forecast error, due to not knowing future values of relevant variables in the models. They suggested the potential usefulness of the alternative of sensitivity analyses, so users of estimates from these models might know if projected costs are sensitive to certain kinds of assumptions. Federal pension plans could be required to use sensitivity analyses in estimating future liability as private plans already must. Concerning the use of recent data on retirement decisions in forecasting pension costs, DOL noted that it may not reduce forecasting error significantly since errors in these models generally relate to retirement behavior occurring years in the future. Third, DOL points out that the National Longitudinal Survey includes repeated interviews with a sample of women now approaching retirement age, which will eventually fill the gap of data on women's retirement that we discuss in our report. Lastly, DOL mentions that the size of a pension plan affects the resources available for cost analyses. We found all these points valuable, though they suggested no specific changes to our text.

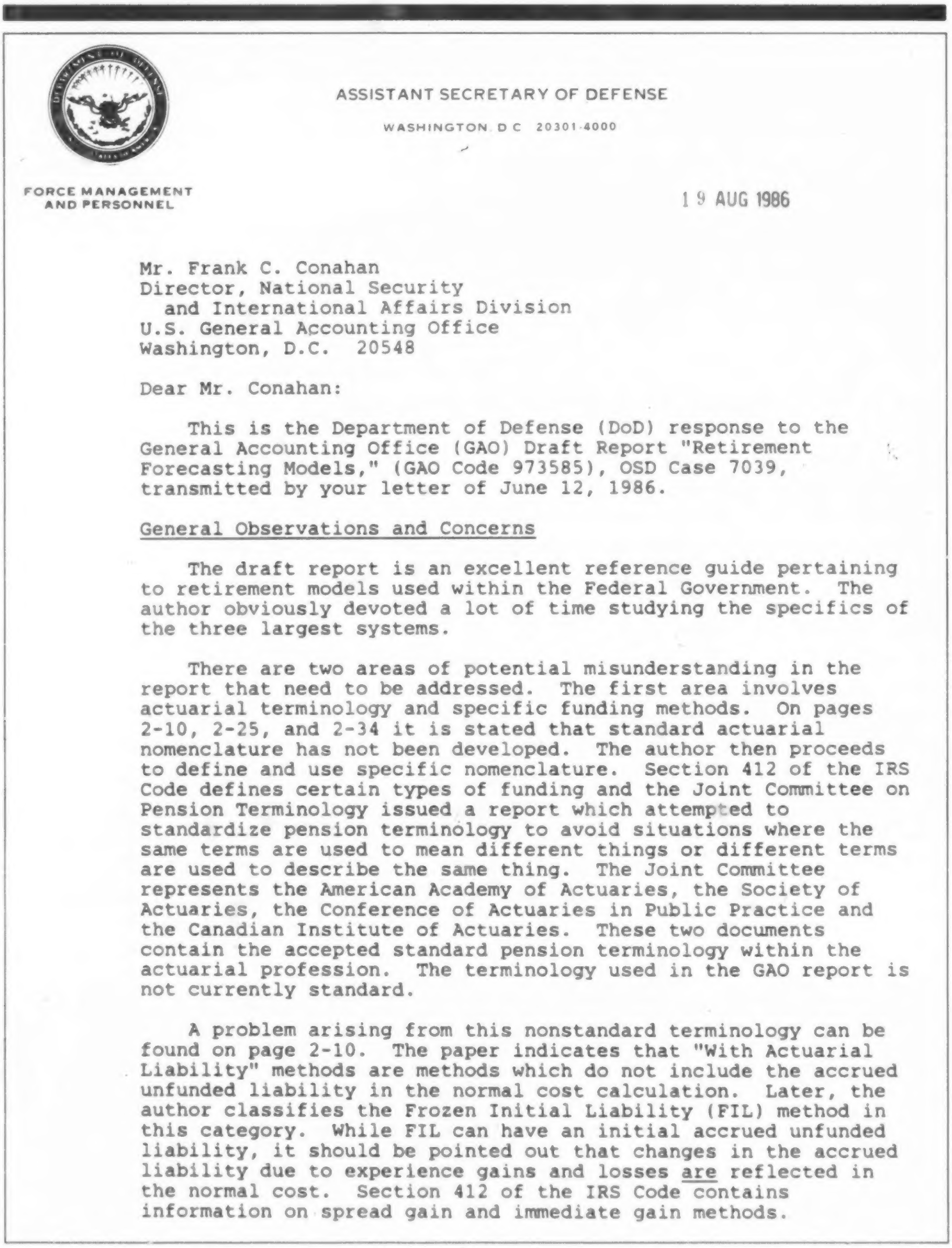
DOL states that our report is critical of retirement decision models for lack of updating and lack of use in forecasting. We did not intend to be critical but simply to describe the models' development and use. Our discussions in chapter 3 shows that, while individual models are not commonly updated, considering them as a class these models are frequently updated and revised. We also noted that some are used for forecasting and some are not, and that several are published with warnings against such use.

DOL states that retirement cost and retirement income models serve different purposes, so that, for example, a major national retirement income simulation may not be useful to analysts concerned with a cost model for a small pension plan covering few people. We agree, and do not believe our report suggests anything else.

Office of Personnel Management (OPM)

OPM found our description and review of the CSRS cost estimate model accurate. They state that over the last few years they have improved and currently have plans for further improvement in the documentation of that model for internal use by actuaries. They made no comments about other aspects of our report.

Comments From the Department of Defense



ASSISTANT SECRETARY OF DEFENSE
WASHINGTON DC 20301-4000

FORCE MANAGEMENT
AND PERSONNEL

19 AUG 1986

Mr. Frank C. Conahan
Director, National Security
and International Affairs Division
U.S. General Accounting Office
Washington, D.C. 20548

Dear Mr. Conahan:

This is the Department of Defense (DoD) response to the General Accounting Office (GAO) Draft Report "Retirement Forecasting Models," (GAO Code 973585), OSD Case 7039, transmitted by your letter of June 12, 1986.

General Observations and Concerns

The draft report is an excellent reference guide pertaining to retirement models used within the Federal Government. The author obviously devoted a lot of time studying the specifics of the three largest systems.

Now pp. 31, 42, and 48.

There are two areas of potential misunderstanding in the report that need to be addressed. The first area involves actuarial terminology and specific funding methods. On pages 2-10, 2-25, and 2-34 it is stated that standard actuarial nomenclature has not been developed. The author then proceeds to define and use specific nomenclature. Section 412 of the IRS Code defines certain types of funding and the Joint Committee on Pension Terminology issued a report which attempted to standardize pension terminology to avoid situations where the same terms are used to mean different things or different terms are used to describe the same thing. The Joint Committee represents the American Academy of Actuaries, the Society of Actuaries, the Conference of Actuaries in Public Practice and the Canadian Institute of Actuaries. These two documents contain the accepted standard pension terminology within the actuarial profession. The terminology used in the GAO report is not currently standard.

Now p. 31.

A problem arising from this nonstandard terminology can be found on page 2-10. The paper indicates that "With Actuarial Liability" methods are methods which do not include the accrued unfunded liability in the normal cost calculation. Later, the author classifies the Frozen Initial Liability (FIL) method in this category. While FIL can have an initial accrued unfunded liability, it should be pointed out that changes in the accrued liability due to experience gains and losses are reflected in the normal cost. Section 412 of the IRS Code contains information on spread gain and immediate gain methods.

Now p. 31.

A concept within the report pertaining to a specific funding method and not relating to the terminology problem can be found on page 2-9. The author explains that accrued benefit methods (by definition) cannot have an actuarial liability. This is not the case. Past service credited prior to funding will generate initial liabilities. Additionally, changes in assumptions, benefits or experience gains and losses will generate additional unfunded liabilities.

The second area of potential misunderstanding pertains to the reasons for actuarial retirement cost forecasting models. The draft report describes the lack of information on validity and forecasting accuracy. It should be stressed in the report that actuarial models are long-term forecasting models which are meant to be accurate in the aggregate over approximately a 100-year period. Assumptions are developed that are valid over time, not in any one year. It is expected that under predictions or over predictions will occur in any one year.

Now pp. 6 and 48.

Pages vii and 2-35 state that forecast accuracy has not been analyzed in the DoD. The DoD produces an annual analysis of all decrement rates used in the projection. Actual-to-expected experiences studies are published each year. Currently, the DoD also calculates detailed actuarial gains and losses in order to adhere to the funding requirements of PL 98-94. These types of annual analyses are good validity indicators. Chapter 95, title 31, U.S.C. requires each Federal pension plan to report annually to Congress. A gain and loss analysis could be added to the requirement. The GAO and the OMB design this reporting format.

Now pp. 35 and 40.

One other minor problem was discovered. Pages 2-16 and 2-22 state that the military model used the OASDI mortality assumptions to construct their own unisex mortality tables. The DoD uses its own military-specific data to calculate mortality tables. In the projection program, the DoD improves these death rates over time by improvement factors developed for the OASDI.

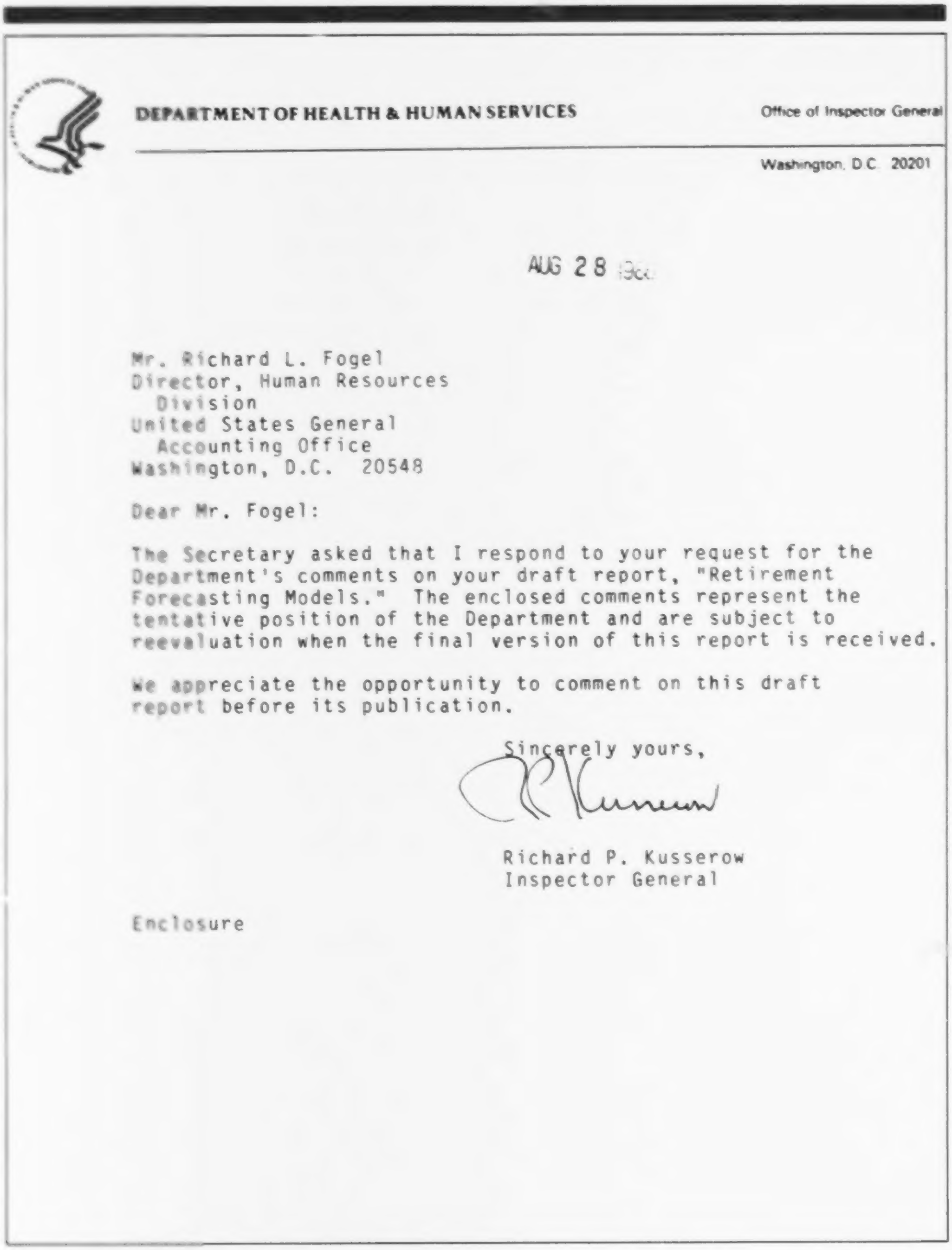
Finally, the DoD has several sophisticated retirement decision models. None of them were mentioned in the text.

Sincerely,

David J. Armor
Principal Deputy

Enclosure

Comments From the Department of Health and Human Services



DEPARTMENT OF HEALTH & HUMAN SERVICES
Office of Inspector General
Washington, D.C. 20201

AUG 28 1986

Mr. Richard L. Fogel
Director, Human Resources
Division
United States General
Accounting Office
Washington, D.C. 20548

Dear Mr. Fogel:

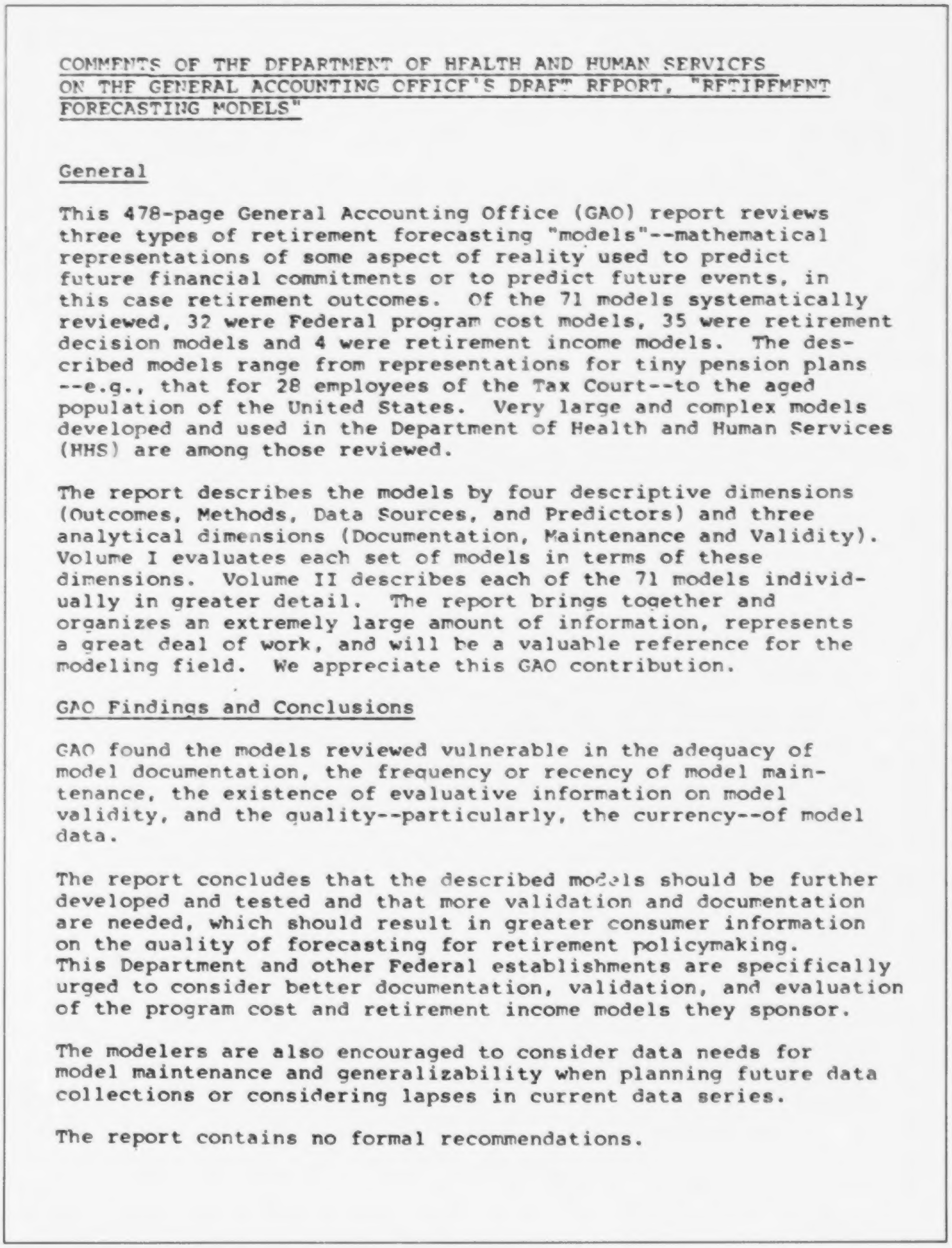
The Secretary asked that I respond to your request for the Department's comments on your draft report, "Retirement Forecasting Models." The enclosed comments represent the tentative position of the Department and are subject to reevaluation when the final version of this report is received.

We appreciate the opportunity to comment on this draft report before its publication.

Sincerely yours,

Richard P. Kusserow
Inspector General

Enclosure



COMMENTS OF THE DEPARTMENT OF HEALTH AND HUMAN SERVICES ON THE GENERAL ACCOUNTING OFFICE'S DRAFT REPORT, "RETIREMENT FORECASTING MODELS"

General

This 478-page General Accounting Office (GAO) report reviews three types of retirement forecasting "models"--mathematical representations of some aspect of reality used to predict future financial commitments or to predict future events, in this case retirement outcomes. Of the 71 models systematically reviewed, 32 were Federal program cost models, 35 were retirement decision models and 4 were retirement income models. The described models range from representations for tiny pension plans --e.g., that for 28 employees of the Tax Court--to the aged population of the United States. Very large and complex models developed and used in the Department of Health and Human Services (HHS) are among those reviewed.

The report describes the models by four descriptive dimensions (Outcomes, Methods, Data Sources, and Predictors) and three analytical dimensions (Documentation, Maintenance and Validity). Volume I evaluates each set of models in terms of these dimensions. Volume II describes each of the 71 models individually in greater detail. The report brings together and organizes an extremely large amount of information, represents a great deal of work, and will be a valuable reference for the modeling field. We appreciate this GAO contribution.

GAO Findings and Conclusions

GAO found the models reviewed vulnerable in the adequacy of model documentation, the frequency or recency of model maintenance, the existence of evaluative information on model validity, and the quality--particularly, the currency--of model data.

The report concludes that the described models should be further developed and tested and that more validation and documentation are needed, which should result in greater consumer information on the quality of forecasting for retirement policymaking. This Department and other Federal establishments are specifically urged to consider better documentation, validation, and evaluation of the program cost and retirement income models they sponsor.

The modelers are also encouraged to consider data needs for model maintenance and generalizability when planning future data collections or considering lapses in current data series.

The report contains no formal recommendations.

Department Comments

We agree that sound retirement-related forecasts, or projections, are important for national policymaking. This is a key responsibility in major HHS-administered programs. We agree that documentation and validation of retirement models are important aspects of model quality. We will describe in the following paragraphs some actions we have taken (mostly since the authors of this report completed their data gathering in 1984) and are taking to update and improve documentation and validation of models we work with. We will also note, in our sequential comments on the chapters of the report, some points the authors overlooked in describing HHS models, but first, some general observations.

The report states that one of its key strengths is accuracy. The comments that we have included should somewhat further improve the report's accuracy. However, it should be noted that accuracy of description does not ensure the imparting of a complete understanding and the ability to form expert opinion on the models. It seems that in an effort to evaluate all of the retirement forecasting models by a standard set of criteria, the purpose of each model's projections is lost. Also, the report does not note the impressive past contributions of models used to assess Federal governmental operations. While it is certainly proper for GAO to note where models have problems with validity, evaluation, and documentation, they do not give the same consideration to the constraints in time and money faced by the developers of these important tools of policy research.

The core issue of the report is the extent of error in the model estimates. The central problem is that in many cases there is simply no known distribution of error around the point estimates provided in the complicated econometric models surveyed in the report. The modelers will claim that they are not really forecasting the future, but making projections based on the conditional assumptions used for the independent variables. There may be thousands of such assumptions required to prepare such a projection.

The projection models themselves are subject to many limitations. Numerous simplifications and approximations are necessarily involved and others are required if the model is to perform responsively. Improvements can be made in this area, and are made regularly, but the potential degree of improvement is minor compared to the effects of unanticipated changes in economic or demographic conditions. Thus, while we certainly concur with GAO's general encouragement to study and improve the models' forecasting ability, we would prefer that the report indicate more clearly the difficulties involved and the limited improvement that is ultimately possible.

Volume I, Chapter 1

Under the "Scope and Methodology" sub heading of this chapter (and elsewhere in both volumes of the report) the authors make a somewhat half-hearted attempt to present the Federal Government role as employer, concerned with the retirement income security of Federal workers and the costs to the Government of providing benefits to Federal employees, as the unifying consideration determining the selection of models for examination. The inventory does not hang together on that basis. Neither the retirement decision models nor the retirement income models center on Federal employee data, although the many small program cost models do. The giant Social Security models and the National Institute on Aging's Macroeconomic-Demographic Model (MDM) apply to practically the whole population. Perhaps the unifying theme should be Federal program responsibilities. It is possible that the same analysis and same standards are not appropriate for simple pension fund cost models and large simulation models.

Chapter 2

Now p. 25.

Now p. 25.

On page 2-3, credit is given to the Social Security Administration's (SSA) Office of the Actuary for making forecasts for the Old Age, Survivors and Disability Insurance (OASDI) program and it is noted on page 2-2 that the primary use of these forecasts is to generate projections for the annual Trustees' Reports. It should be noted, however, that the ultimate responsibility for the projections shown in the Trustees' Reports rests with the trustees themselves.

Now p. 37.

On page 2-17, it is mentioned that a higher inflation rate is a more conservative assumption because it makes the normal cost of a pension plan higher. However, for the OASDI program, a higher inflation rate makes the long-range average cost rate lower. This is an important difference, especially in light of the last paragraph on page 2-19, which is very misleading.

Now p. 38.

On page 2-19, although the OASDI model generally uses economic and demographic assumptions that provide a smooth trend toward an ultimate value, the intermediate years are not determined by interpolation. Also, although by the year 2010 all assumptions have reached their ultimate values, as mentioned in the first paragraph on page 2-19, most of them are actually reached in an earlier year.

Now pp. 42 and 48.
See also vol. 2, p. 20.

On page 2-26 and again on 2-34 and in Volume II at I-20, the report states that OASDI model documentation is not complete. GAO may want to refer to two publications which summarize the entire long-range model: the annual OASDI Trustees' Report itself, and Actuarial Study Number 91 published in April 1984. Descriptions of the short-range model are included in the OASDI Trustees Report, beginning with the 1986 report.

Now p. 16. See also vol. 2, p. 21.

On page 2-32 and in Volume II at page I-21, characterization of the "...five-year cost estimates that differ from actual experience by as much as 40 percent of annual benefit payments" is very misleading, because the cost estimate referred to is not specified. This estimate is of the fifth-year trust fund balance. Because the trust fund balance is the difference between two large numbers (income and outgo) and is cumulative through time, the fifth-year trust fund balance is very sensitive even to small differences between actual and projected experience. Because cost estimates are prepared as a percentage of payroll in order to compare costs with legislated tax rates, this is a much more appropriate model output to judge model validity. Members of Congress should not be left with the impression that they may have to change tax rates or benefit rates by 40 percent within 5 years.

Now p. 48.

Contrary to various statements made in the report (e.g. on page 2-35), there is a regular, ongoing evaluation of "forecast accuracy" of the OASDI short-range model that has been in existence for quite some time. It is a comparison of the most recent actual year of experience (for tax income and benefit payments) with the estimated amounts from the two preceding Trustees' Reports. It is shown in each year's Trustees' Report (as table 4 in the current report) and is accompanied by a brief explanation of the comparison and any relevant issues.

In general, the OASDI projections are subject to three sources of inaccuracies: economic and demographic experience that does not follow the assumptions, legislative and regulatory changes in the OASDI program, and limitations of the projection models themselves. The first of these sources is often the most problematical. It is not now possible to set most key assumptions with any degree of certainty and is unlikely ever to become possible. While it is of interest to track actual economic experience versus the range of assumptions (as has been done for a number of years in SSA), past "successes or failures" provide very little guidance concerning future accuracy.

With respect to changes in the OASDI program, a proper evaluation of a model's accuracy can only be made by factoring out any such changes that have occurred during the projection period. We are often hard-pressed to develop the necessary modifications to our models to reflect new legislation in a timely manner; maintenance of the older version(s) in addition, to allow such a comparison, has generally not been feasible. Finally, the projections themselves can be improved, and we do that, but the impact of the possible improvements is small compared to the impact of changes in demographic and economic conditions that are not predictable.

Chapter 3

The evaluation presented in Volume I would have been far more useful if the critique of models, as defined in the study, were concentrated on the purpose of the model rather than some arbitrary standard. The models on the retirement decision, for example, were, in part, evaluated on the basis of whether or not they were updated periodically. These models were not designed to be updated. Instead, these 30 or so individual studies attempted to explain the determinants of the decision to retire. Some of them were expansions or improvements on earlier studies of retirement behavior; others tested alternative specifications. The GAO report does not compare the results from these studies with each other. A discussion of the appropriateness of the determinants (predictors) used in the model, specification of the variables, statistical significance, and other statistical properties of the model would have been useful and appropriate. In addition, the discussion of the retirement decision models would benefit greatly from an evaluation of whether the studies yielded any consistent (or inconsistent) results with respect to the importance of specific determinants of the decision to retire. If, for example, availability of private pensions was a variable that was statistically significant in most of the studies irrespective of the time period, specific formulation, or data base, then this variable would appear to be an important determinant of the decision to retire. But nowhere does the GAO report contain any of this kind of comparative analysis of the structure of empirical results of any of the models reviewed in the study.

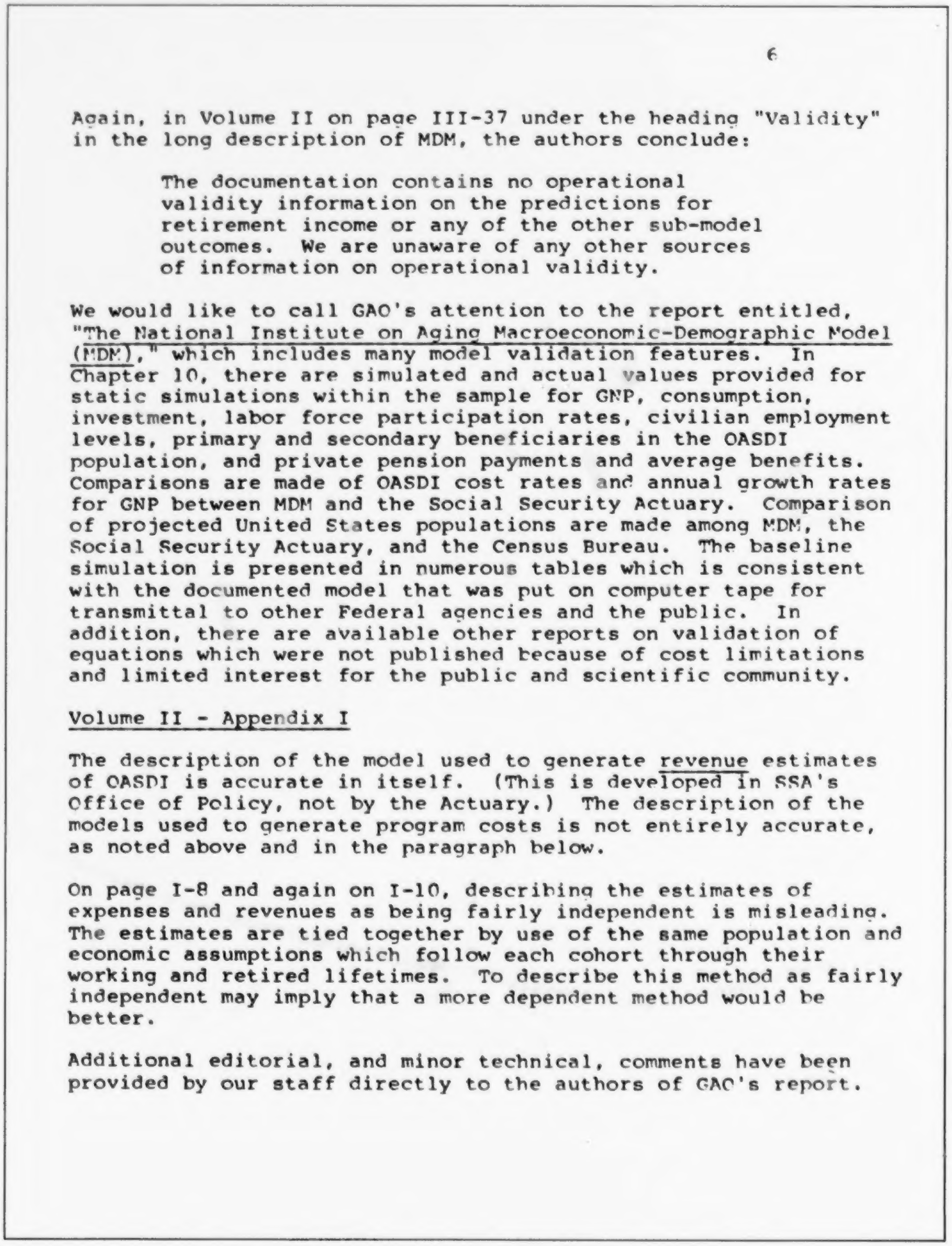
The GAO report does highlight the need to provide current--particularly survey--data so that models can be estimated based on recent experience or behavior patterns of current cohorts.

Chapter 4

Now p. 84.

On page 4-16 the authors state:

> There is little published information on the operational validity of these [retirement income] models....No information is available... on the potential for forecast error in final outcomes...Developers reported to us that they monitor the accuracy of their assumptions, calculate validity statistics on estimated equations and perform sensitivity analyses. However, the results of these analyses are not routinely published.



See vol. 2, p. 150.

Again, in Volume II on page III-37 under the heading "Validity" in the long description of MDM, the authors conclude:

> The documentation contains no operational validity information on the predictions for retirement income or any of the other sub-model outcomes. We are unaware of any other sources of information on operational validity.

We would like to call GAO's attention to the report entitled, "The National Institute on Aging Macroeconomic-Demographic Model (MDM)," which includes many model validation features. In Chapter 10, there are simulated and actual values provided for static simulations within the sample for GNP, consumption, investment, labor force participation rates, civilian employment levels, primary and secondary beneficiaries in the OASDI population, and private pension payments and average benefits. Comparisons are made of OASDI cost rates and annual growth rates for GNP between MDM and the Social Security Actuary. Comparison of projected United States populations are made among MDM, the Social Security Actuary, and the Census Bureau. The baseline simulation is presented in numerous tables which is consistent with the documented model that was put on computer tape for transmittal to other Federal agencies and the public. In addition, there are available other reports on validation of equations which were not published because of cost limitations and limited interest for the public and scientific community.

Now appendix II.

Volume II - Appendix I

The description of the model used to generate revenue estimates of OASDI is accurate in itself. (This is developed in SSA's Office of Policy, not by the Actuary.) The description of the models used to generate program costs is not entirely accurate, as noted above and in the paragraph below.

See vol. 2, pp. 14 and 15.

On page I-8 and again on I-10, describing the estimates of expenses and revenues as being fairly independent is misleading. The estimates are tied together by use of the same population and economic assumptions which follow each cohort through their working and retired lifetimes. To describe this method as fairly independent may imply that a more dependent method would be better.

Additional editorial, and minor technical, comments have been provided by our staff directly to the authors of GAO's report.

Comments From the Department of Labor

THE UNDER SECRETARY OF LABOR
WASHINGTON, D.C.
20210

July 15, 1986

Mr. Richard L. Fogel
Director
Human Resources Division
United States General Accounting
  Office
441 G Street, N.W.
Washington, D.C. 20548

Dear Mr. Fogel:

Thank you for the opportunity to comment on GAO's draft proposed report "Retirement Forecasting Models." The report is an impressive compendium of the general characteristics and state of documentation and validation of the over 70 models that are used to forecast federal retirement program costs and analyze civilian retirement behavior and retirement income. The study should provide a considerable service to researchers and practitioners in the retirement area, and will help to emphasize the need to collect the data that is so critical to the development and use of this important set of models.

The Department's specific comments and suggestions for possible improvements in the report are enclosed. We hope our observations will serve to further the discussion of the appropriate methods to be used in estimating the future costs of federal pension programs. The GAO has provided an excellent starting point for that discussion.

Thank you again for the opportunity to comment. We look forward to your final report.

Sincerely,

Dennis E. Whitfield

DENNIS E. WHITFIELD

Enclosures

DEW:hlr

DEPARTMENT OF LABOR COMMENTS ON DRAFT GAO REPORT:
RETIREMENT FORECASTING MODELS

Retirement Program Cost Models

One of the primary focuses of this portion of the draft report is the accuracy of the cost estimates produced by models of federal retirement programs. The report is critical of the users of these actuarial models because they do not calculate forecasting errors for their estimates of future pension costs. GAO is correct in asserting that federal pension programs generally follow the accepted actuarial practice for private plans, where forecast error also is not calculated. However, for reasons outlined below, the practice of not calculating forecasting error may be justified for this type of model.

Forecasting error can be broadly categorized into two types. The first is error in the specification of the mathematical model used for forecasting. An example of this type of error is not including all the relevant variables in the model. We doubt that the actuaries for federal pension programs are guilty of this error. The basic variables affecting pension costs -- primarily job separation, mortality, wage earnings, the benefit formulas, and investment earnings -- are well established.

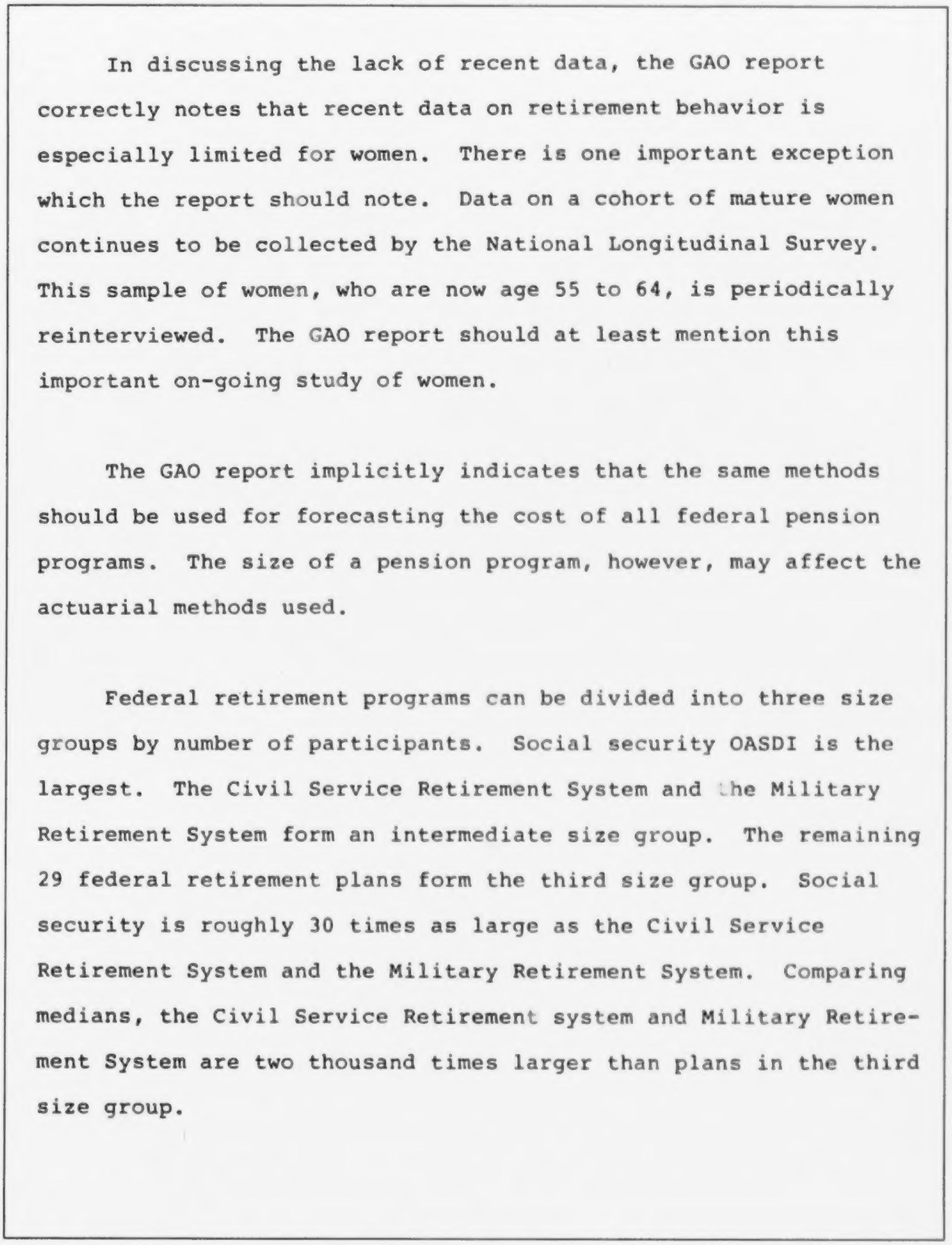
The second type of error is error due to not knowing the future values of relevant variables. For example, the future number of pensioners in a federal retirement program is not known with certainty nor is the exact distribution of the ages at which future pensioners will retire. It is difficult, if not impossible, to develop meaningful measures to gauge the forecasting error attributable to this second error source. However, because of the second type error, sensitivity analysis should be done to assess the extent to which small errors in the projected values of some variables may greatly affect the pension cost forecasts. Those who use the estimates produced by these models should know if the projected costs are highly sensitive to certain types of assumptions.

A general approach that could be suggested by the GAO report is that federal pension plans be subject to the same accounting rules as are private plans, as determined by the Financial Accounting Standards Board (FASB). FASB recently required that private plans do some sensitivity analysis in determining their future pension liability. This suggestion is in the spirit of other suggestions made in the GAO report.

The GAO report is also critical of the lack of recent data on retirement behavior. With regard to actuarial cost models, the second type error is not likely to be affected very much by the availability of recent data, since this type or error generally relates to retirement behavior occurring many years in the future. (The first type of error is also not affected by the lack of recent data, since, as already indicated, the basic predictors relevant to actuarial models of pension costs are well established.)

Recent data is most important for cost models which are not of the standard actuarial type. With regard to Federal programs, this includes models which forecast Social Security OASDI program costs. However, even here, having very recent data on labor force participation, age of retirement and other relevant variables provides no assurance that predictions of retirement behavior many years in the future will be more accurate.

As the GAO report correctly points out, recent data is very important for studying the factors which affect the decision to retire. However, retirement decision models are most useful not in forecasting program costs but rather in analyzing changes in costs in response to a change in policy. The Department suggests that the GAO spell out more clearly in its final report that current data is needed to better understand the impact of changes in policy but would not necessarily reduce the forecasting error in the actuarial models used for making the standard annual forecasts of federal program costs.



In discussing the lack of recent data, the GAO report correctly notes that recent data on retirement behavior is especially limited for women. There is one important exception which the report should note. Data on a cohort of mature women continues to be collected by the National Longitudinal Survey. This sample of women, who are now age 55 to 64, is periodically reinterviewed. The GAO report should at least mention this important on-going study of women.

The GAO report implicitly indicates that the same methods should be used for forecasting the cost of all federal pension programs. The size of a pension program, however, may affect the actuarial methods used.

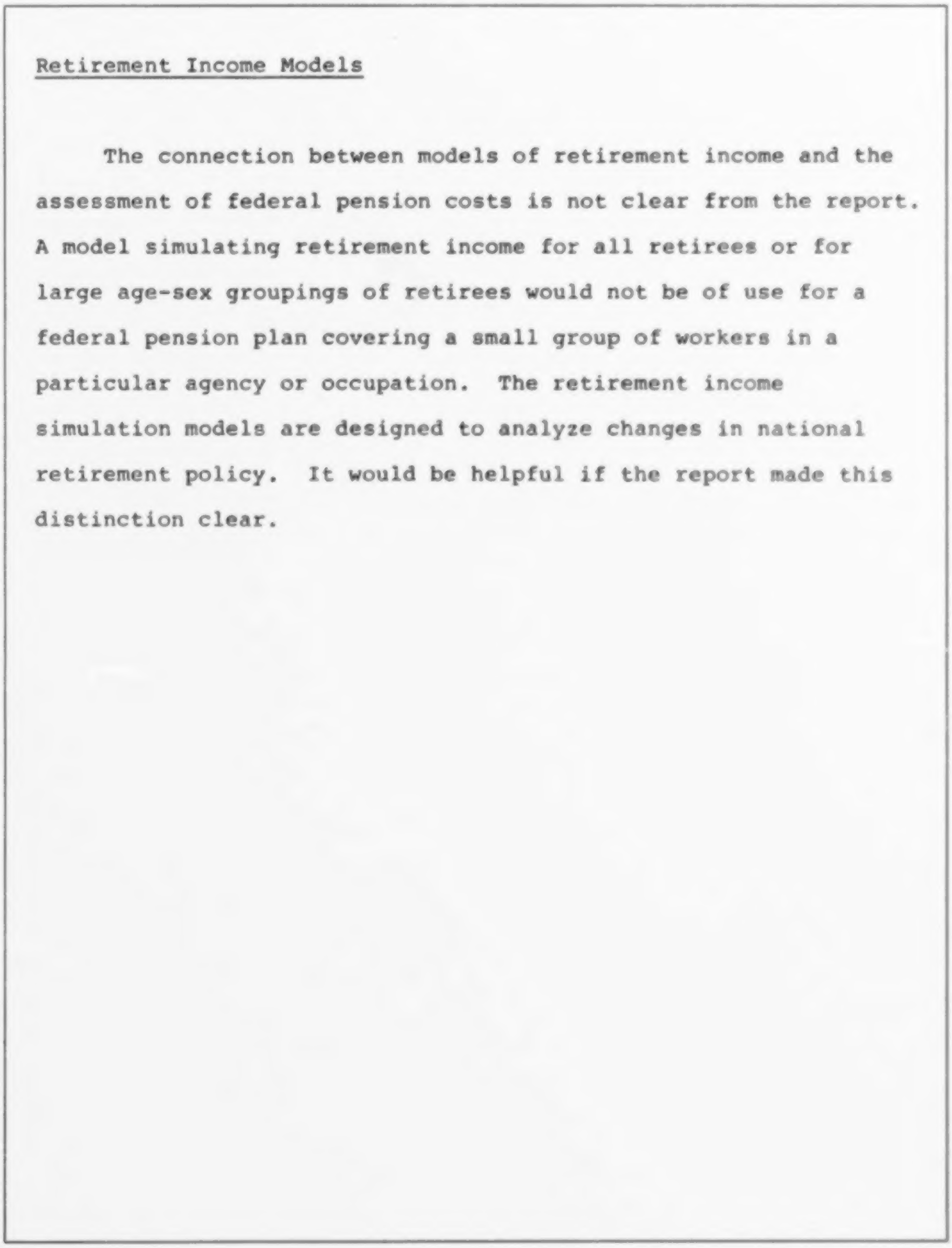
Federal retirement programs can be divided into three size groups by number of participants. Social security OASDI is the largest. The Civil Service Retirement System and the Military Retirement System form an intermediate size group. The remaining 29 federal retirement plans form the third size group. Social security is roughly 30 times as large as the Civil Service Retirement System and the Military Retirement System. Comparing medians, the Civil Service Retirement system and Military Retirement System are two thousand times larger than plans in the third size group.

Because of the large differences in plan size, the amount of resources used to determine future pension liabilities and, hence, the sophistication of the effort would be expected to differ considerably across the three plan size categories. In the Department's opinion, the final GAO report should not imply that even very small Federal pension programs should use highly sophisticated and very costly models in forecasting costs.

Retirement Decision Models

The GAO report is critical of retirement decision models because they are not updated and because they are generally not used by their developers for forecasting. Retirement decision models are rigorous scientific models that are developed to advance knowledge about individual behavior. Although these models can and should be used to provide useful information for forecasters, forecasting is not their main purpose. In any case, updating retirement decision models will generally have little affect on either two types of forecasting error discussed above. The report should make this point clear.

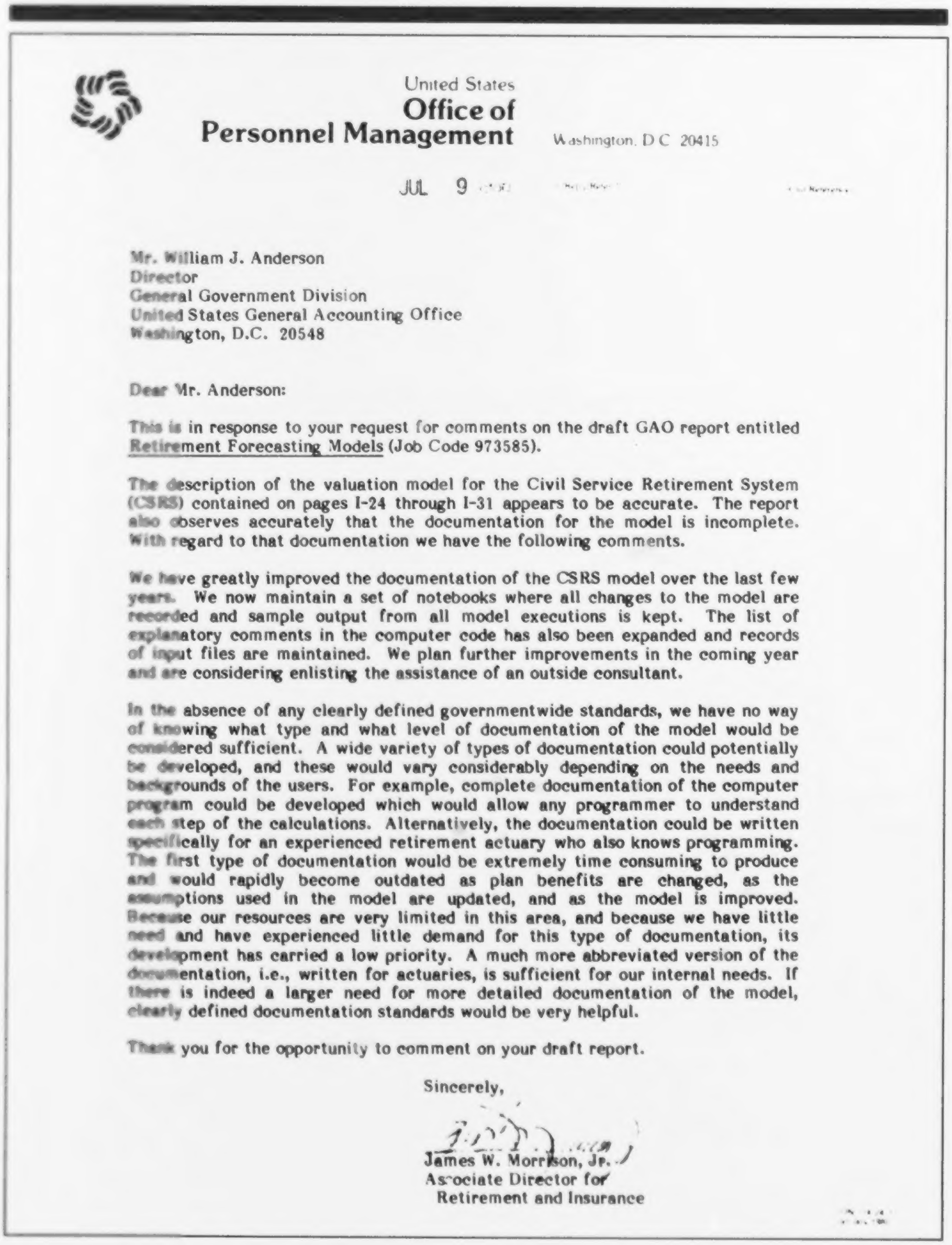
The Department does agree with the report's more general point with regard to these models. Recent data is needed to continue to test and develop these models, which are critical to analyzing the impacts of policy changes.



Retirement Income Models

The connection between models of retirement income and the assessment of federal pension costs is not clear from the report. A model simulating retirement income for all retirees or for large age-sex groupings of retirees would not be of use for a federal pension plan covering a small group of workers in a particular agency or occupation. The retirement income simulation models are designed to analyze changes in national retirement policy. It would be helpful if the report made this distinction clear.

Comments From the Office of Personnel Management



United States
**Office of
Personnel Management**
Washington, D C 20415

JUL 9 [illegible]

Mr. William J. Anderson
Director
General Government Division
United States General Accounting Office
Washington, D.C. 20548

Dear Mr. Anderson:

This is in response to your request for comments on the draft GAO report entitled Retirement Forecasting Models (Job Code 973585).

See vol. 2, pp. 23-27.

The description of the valuation model for the Civil Service Retirement System (CSRS) contained on pages I-24 through I-31 appears to be accurate. The report also observes accurately that the documentation for the model is incomplete. With regard to that documentation we have the following comments.

We have greatly improved the documentation of the CSRS model over the last few years. We now maintain a set of notebooks where all changes to the model are recorded and sample output from all model executions is kept. The list of explanatory comments in the computer code has also been expanded and records of input files are maintained. We plan further improvements in the coming year and are considering enlisting the assistance of an outside consultant.

In the absence of any clearly defined governmentwide standards, we have no way of knowing what type and what level of documentation of the model would be considered sufficient. A wide variety of types of documentation could potentially be developed, and these would vary considerably depending on the needs and backgrounds of the users. For example, complete documentation of the computer program could be developed which would allow any programmer to understand each step of the calculations. Alternatively, the documentation could be written specifically for an experienced retirement actuary who also knows programming. The first type of documentation would be extremely time consuming to produce and would rapidly become outdated as plan benefits are changed, as the assumptions used in the model are updated, and as the model is improved. Because our resources are very limited in this area, and because we have little need and have experienced little demand for this type of documentation, its development has carried a low priority. A much more abbreviated version of the documentation, i.e., written for actuaries, is sufficient for our internal needs. If there is indeed a larger need for more detailed documentation of the model, clearly defined documentation standards would be very helpful.

Thank you for the opportunity to comment on your draft report.

Sincerely,

James W. Morrison, Jr.
Associate Director for
Retirement and Insurance

References

ANDERSON, J. M. Modeling Analysis of Retirement Income Policy: Background and Overview. Washington, D.C.: Employee Benefit Research Institute, 1980.

ASCHER, W. Forecasting: An Appraised for Policy-Makers and Planners. Baltimore, Md.: Johns Hopkins University Press, 1978.

BARTLETT, D. K., III, and J. A. Applebaum. "Economic Forecasting: Effect of Errors on OASDI Fund Ratios." Social Security Bulletin, 45:1 (1982), 9-14.

DANZIGER, S., R. Haveman, and R. Plotnick. "How Income Transfer Programs Affect Work, Savings, and the Income Distribution: A Critical Review." Journal of Economic Literature, 19 (1981), 975-1028.

FEDERAL Old-Age and Survivors Insurance and Disability Insurance Trust Funds. 1983 Annual Report-Federal Old-Age and Survivors Insurance and Disability Insurance Trust Fund. Washington, D.C.: U.S. Government Printing Office, 1983.

FIELDS, G. S. and O. S. Mitchell. "Pensions and the Age of Retirement." Cornell University, Ithaca, N.Y., 1983.

GAO (U.S. General Accounting Office). Food and Agriculture Models for Policy Analysis, CED-77-87. Washington, D.C.: July 13, 1977a.

———. An Evaluation of the Use of the Transfer Income Model— TRIM— to Analyze Welfare Programs, PAD-78-14. Washington, D.C.: November 25, 1977b.

———. Models and their Role in GAO, PAD-78-84. Washington, D.C.: October 1978a.

———. Need for Overall Policy Analysis and Coordinated Management of Federal Retirement Systems, FPCD-78-49. Washington, D.C.: December 29, 1978b.

———. Guidelines for Model Evaluation, PAD-79-17. Washington, D.C.: January 1979.

———. How Much Does the Federal Government Spend on Economic Modeling Activities? PAD-82-26. Washington, D.C.: February 16, 1982a.

GAO. The Results of Our Initial Review of the Financial and Actuarial Statements Submitted Pursuant to Public Law 95-595, AFMD-82-67. Washington, D.C.: May 21, 1982b.

——. Data from the Retirement History Survey, IPE-82-5. Washington, D.C.: July 6, 1982c.

——. Inadequate Internal Controls Affect Quality and Reliability of the Civil Service Retirement System's Annual Report, AFMD-83-3. Washington, D.C.: October 22, 1982d.

——. Review of Fiscal Year 1982 Financial Transactions of the Foreign Service Retirement and Disability System, B-164292. Washington, D.C.: August 18, 1983a.

——. Social Security Actuarial Projections, HRD-83-92. Washington, D.C.: September 30, 1983b.

——. Toward Informed Decision-Making on Retirement Policy: A Portfolio of Studies Related to the Retirement Decision. PEMD Special Study 2. Washington, D.C.: March 1984.

——. Summary of 1983 Federal Pension Plan Information, AFMD-85-69. Washington, D.C.: September 26, 1985.

——. Social Security: Past Projections and Future Financing Concerns, HRD-86-22. Washington, D.C.: March 11, 1986.

GASS, S. I. "Evaluation of Complex Models." University of Maryland, College Park, Md., 1976.

GUSTMAN, A. L. "Modeling Individuals' Behavior: Evaluation of a Policymaker's Tool." Journal of Policy Analysis and Management, 3:2 (1984), 191-205.

——, and T. L. Steinmeyer. "Minimum Hours Constraints and Retirement Behavior." Contemporary Policy Issues, No. 3 (April 1983), 77-91.

HAVEMAN, R. H., and J. M. Lacker. "Discrepancies in Projecting Future Public and Private Pension Benefits: A Comparison and Critique of Two Micro-data Simulation Models." University of Wisconsin, Madison, Wisc., February 1984.

JOINT COMMITTEE ON PENSION TERMINOLOGY. Pension Terminology: Final Report by the Joint Committee. [n.p.], July 31, 1981.

LIGHT, P. C. "Social Security and the Politics of Assumptions." Paper presented at the 1983 American Political Science Association Meeting, Chicago, 1983.

MADDALA, G. S. Limited-dependent and Qualitative Variables in Econometrics. New York: Cambridge University Press, 1983.

McLEOD, J. "Simulation: From Art to Science for Society." Simulation Today, No. 20 (December 1973). (As cited in GAO, 1979.)

McGILL, D. M. Fundamentals of Private Pensions (4th Edition). Homewood, Ill.: Richard D. Irwin, Inc., 1979.

MODIGLIANI, F., and R. Brumberg. "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data." In K. Kurihara (ed.), Post-Keynesian Economics. London: Allen and Unwin, 1955.

MYERS, R. J. "Actual Costs of the Social Security System Over the Years Compared with 1935 Estimates." Social Security Bulletin, 45:3 (1982), 13-15.

Glossary

Actuarial Assumptions	A prediction of future conditions affecting pension cost; for example, mortality rate, employee turnover, compensation levels, investment earnings, etc.
Actuarial Cost Method	A procedure which uses actuarial assumptions to measure the present value of future pension benefits and pension fund administrative expenses and which allocates the cost of such benefits and expenses to time periods.
Actuarial Gain or Loss	A measure of the difference between a plan's actual experience and that expected based on actuarial assumptions.
Actuarial Liability	The portion, as determined by the actuarial cost method in use, of the present value of pension benefits and expenses which is not provided for by future normal costs.
Data Sources	The basic information generated externally that a model processes in making a forecast.
Demographic Assumptions	Assumptions which are concerned with the status of the participant population, such as retirement rates, disability rates, and mortality rates. (See also "Economic assumptions.")
Economic Assumptions	Assumptions which are concerned with economic factors, such as future expected inflation, interest rates, and wage increases.
Forecast Error	A measure of the difference between actual outcomes and their forecast values.
Life Cycle Theory	An economic theory that individuals make decisions based upon an evaluation of their current economic status and their expected future economic status.

Macrosimulation	As used in this report, application of a model that represents the functioning of a system at a group or aggregate level.
Methods	The techniques used to implement a model.
Microsimulation	Application of a model that represents the functioning of a system at the individual or household level.
Normal Cost	The portion, as determined by the actuarial cost method in use, of the present value of pension benefits which is allocated to a valuation year.
Outcomes	The specific results that a model produces.
Pension Plan Participant	An employee, former employee, or beneficiary who may become eligible to receive, or is receiving, benefits under a pension plan as a result of credited service.
Predictors	Factors used to describe different aspects of a system being modeled and to forecast outcomes of that system. Variation in the values for the predictors produces variation in forecasted outcomes. Predictors are often referred to as determinants of those outcomes.
Predictor Value	The particular numerical quantity assigned to a predictor.
Present Value	The current worth of an amount or series of amounts payable or receivable in the future, determined by discounting the future amount or amounts at a predetermined rate of interest.
Unfunded Actuarial Liability	The excess of the actuarial liability, under the actuarial cost method in use, over the value of the assets of a plan.

Requests for copies of GAO reports should be sent to:

U.S. General Accounting Office
Post Office Box 6015
Gaithersburg, Maryland 20877

Telephone 202-275-6241

The first five copies of each report are free. Additional copies are $2.00 each.

There is a 25% discount on orders for 100 or more copies mailed to a single address.

Orders must be prepaid by cash or by check or money order made out to the Superintendent of Documents.